Universidade do Minho
Escola de Engenharia

César Manuel Ferreira Quintas

Disponibilidade, Desempenho e Confiança em Sistemas de Bases de Dados Hospitalares - Suporte à Decisão Inteligente para Prevenção de Falhas

maio de 2017

Disponibilidade, Desempenho e Confiança em Sistemas de Bases de Dados Hospitalares - Suporte à Decisão Inteligente para Prevenção de Falhas
César Manuel Ferreira Quintas
UMinho | 2017

Universidade do Minho
Escola de Engenharia

César Manuel Ferreira Quintas

Disponibilidade, Desempenho e Confiança em Sistemas de Bases de Dados Hospitalares - Suporte à Decisão Inteligente para Prevenção de Falhas

Tese de Doutoramento
Engenharia Biomédica

Trabalho efectuado sob a orientação do
Professor Doutor José Manuel Ferreira Machado
Professor Doutor Manuel Filipe Vieira Torres dos Santos

maio de 2017

DECLARAÇÃO DE INTEGRIDADE

Declaro ter atuado com integridade na elaboração da presente tese. Confirmo que em todo o trabalho conducente à sua elaboração não recorri à prática de plágio ou a qualquer forma de falsificação de resultados. Mais declaro que tomei conhecimento integral do Código de Conduta Ética de Universidade do Minho.

Universidade do Minho, *4* de *Maio* de *2017*

Nome completo, *César Manuel Ferreira Quintas*

Assinatura

Agradecimentos

Só foi possível realizar esta tese de doutoramento com a colaboração estreita de várias pessoas, tais como:os meus familiares, professores, colegas de trabalho e amigos. Realço a instituição onde trabalho, onde foi possível aproveitar as ideias inovadoras, que contribuíram para o desenvolvimento deste trabalho.

Aos meus orientadores, o Professor Doutor Filipe Santos e Professor Doutor José Machado, o meu profundo agradecimento pelas suas ideias, críticas pertinentes, ensinamentos, pela disponibilidade, empenho e dedicação neste trabalho.

Por fim, à minha mulher Dominique, ao meu filho Diogo, à minha mãe Elvira e meu pai Júlio, pelo apoio incansável e incondicional nesta "maratona".

Resumo

Nos ambientes hospitalares, as tecnologias e os sistemas de informação suportam a atividade dos profissionais que atuam na prestação de cuidados de saúde, devendo proporcionar um acesso seguro e consistente, confidencialidade, eficiência e disponibilidade permanente. Por outro lado, devem mimetizar os processos de registo e difusão da informação mantendo um elevado grau de desempenho.

Nestes contextos, os Sistemas de Gestão Bases de Dados (SGBD) assumem-se como sistemas críticos consumindo grandes quantidades de recursos de armazenamento, de processamento e de comunicação e não podendo estar sujeitos a falhas.

Apesar da existência de ferramentas tecnológicas para a monitorização e a afinação das configurações dos SGDBs e do avanço tecnológico, com versões mais estáveis, mais seguras e com maior desempenho constata-se que os problemas e dificuldades continuam a surgir.

Estes aspectos constituem uma oportunidade de investigação no sentido de se procurar definir modelos para a previsão de eventos, situações e formas de atuação, com alguma antecedência, tais como:

- Prever a taxa de utilização de um sistema e do volume de armazenamento necessário;
- Prever a ocorrência de um evento crítico que implique a paragem do SGDB;
- Prever com antecedência os recursos necessários para resolver um problema.

A investigação na área clínica há muito que tem produzido modelos de intervenção em situações críticas permitindo alocar recursos e atuar em conformidade (e.g. MEWS). Seguindo de perto estes avanços, este trabalho visa o estudo, modelação e implementação de um modelo de antecipação e de intervenção em SGBDs e a sua materialização no contexto de um sistema de suporte à decisão inteligente.

Abstract

In healthcare environments, technologies and information systems support the activities of professionals working in health care and must provide a safe and consistent access, confidentiality, efficiency and continuous availability. On the other hand, should reduce the registration process and dissemination of information while maintaining a high performance degree.

In these contexts, the Database Management Systems (DMS) are assumed to be critical systems consuming large amounts of storage resources, processing and communication and can not be subject to failures.

Despite the existence of technological tools for monitoring and tuning the settings of DMS and technological advances, more stable versions, safer and higher performance notes that the problems and difficulties continue to arise. These constitute a research opportunity in order to seek to define models for the prediction of events, situations and ways of action, such as:

- Predict the utilization rate of a system and the required storage volume;
- Provide for the occurrence of a critical event that causes the stop of DMS;
- Plan in advance the resources needed to solve a problem.

Research in the clinical area has long produced models of intervention in critical situations allowing allocate resources and act accordingly (e.g. MEWS). Following these developments, this work aims to study, modeling and implementation of a model of anticipation and intervention in DBMS and its materialization in the context of a support system for intelligent decision.

Índice

Acrónimos

ACSS Administração Central do Sistema de Saúde

ADL *Archetype Definition Language*

AIDA Agência para a Integração, Difusão e Arquivo

AM *Archetype Model*

ANSI *American National Standards Institute*

ARRA *American Recovery and Reinvestment Act*

AUI *Atom Unique Identifier*

CAP *College of American Pathologists*

CC Ciências da Computação

CUI *Concept Unique Identifier*

CEN Comissão Europeia de Normalização

DICOM *Digital Imaging and Communications in Medicine*

DR *Design Research*

EC Estudo de Caso

FSN *Fully Specified Name*

EMRAM *Electronic Medical Record Adoption Model*

EUA *Estados Unidos da América*

GIA Grupo de Inteligência Artificial

HIMSS *Healthcare Information Managment Systems Society*

HL7 *Health Level Seven*

HL7 CDA *HL7 Clinical Document Arquitecture*

HL7 RIM *HL7 Reference Information Model*

IA Inteligência Artificial

ICD *International Classification of Diseases*

ICD9 *ICD 9th Revision*

ICD9-CM *ICD9 Clinical Modification*

ICD10 *ICD 10th Revision*

ICD10-CM *ICD10 Clinical Modification*

ICN *International Council of Nurses*

ICNP *International Classification for Nursing Practice*

IGIF Administra ção Central do Sistema de Saúde

IHTSDO *Internetional Health Terminology Standards Development Organization*

IM Informática Médica

IOM *Institute of Medicine*

ISO *International Standards Organization*

IOS EN 13606 *International Standards Organization European Norm 13606*

LOINC *Logical Observation Identifiers Names and Codes*

LUI *Lexical Unique Identifier*

MCDT Meio Complementar de Diagónico e Terapêutico

MSA Módulo *Stand-Alone*

NCHS *National Center for Health Statistics*

NLM *National Library of Medicine*

PACS *Picture Archive and Communication System*

PAQC *Patient Access to Quality Care*

PC Processo Clínico

PCE Processo Clínico Eletrónico

PoC Prova de Conceito

POMR *Problem-oriented medical record*

PCP Processo Clínico em Papel

SAM Sistema de Apoio Médico

SAPE Sistema de Apoio às Práticas de Enfermagem

SINUS Sistema de Informação para Unidades de Saúde

SI Sistemas de Informação

SIH Sistemas de Informação Hospitalar

SM *Service Model*

SMS *Short Message Service*

SNOMED CT *Systematized Nomenclature of Medicine Clinical Terms*

SNOMED RT *SNOMED Reference Terminology*

SNS Serviço Nacional de Saúde

SOA *Service-Oriented Architectures*

SONHO Sistema Integrado de Informação Hospitalar

SPMS Serviços Partilhados do Ministério da Saúde

SSD Sistemas de Suporte à Desição

SUI *String Unique Identifier*

RELMA *Regenstrief LOINC Mapping Assistant*

RM *Reference Model*

TCP/IP *Transmission Control Protocol/Internet Protocol*

TI Tecnologias da Informação

UDDI *Universal Description, Discovery and Integration*

UMLS *Unified Medical Language System*

UMLS M *UMLS Metathesaurus*

UMLS SL *UMLS Specialist Lexicon*

UMLS SN *UMLS Semantic Network*

WHO *World Health Organization*

WIMP *Window, Icon, Menu and Pointing device*

WS *Web Service*

WSDL *Web Services Description Language*

XML *eXtensible Markup Language*

Lista de Tabelas

Lista de Figuras

Capítulo 1

Introdução

1.1 Introdução aos Conceitos

A constante evolução das tecnologias da computação conduziu ao processo massivo de informatização dos serviços nas organizações das mais diversas áreas (por exemplo, empresas, universidades, bibliotecas e hospitais). Ao mesmo tempo, a importância da informação para as organizações aumentou consideravelmente, sendo que hoje em dia, o sucesso de uma organização está relacionado com a forma com que esta gere a sua informação. Esta panóplia atual atribuiu às bases de dados um papel fundamental em qualquer tipo de organização. A utilização de base de dados torna possível proceder ao armazenamento e à gestão da gigantesca quantidade de dados presente numa organização, possibilitando um acesso simples e eficaz a toda a informação. Para as organizações, estas vantagens tornam as bases de dados imprescindíveis na realização das tarefas do dia-a-dia [Silva, 2012].

Esta importância das bases de dados torna-se ainda mais elevada nas organizações que prestam serviços críticos, como é o caso das organizações de prestação de serviços de saúde. Na maioria destas instituições já se encontram implementadas bases de dados que contêm as informações administrativas e clínicas relativas aos utentes. Para além disso, as informações armazenadas são usadas como suporte às decisões e atos clínicos, às decisões e processos administrativos e a um conjunto de aplicações responsáveis por gerir, disponibilizar e promover a partilha da informação. As bases de dados assumem assim um papel vital nas instituições hospitalares [Silva, 2012].

Bases de Dados

As bases de dados são os sistemas que garantem a manutenção, o armazenamento organizado dos dados e a segurança dos mesmos [Tereso and Bernardino, 2011] [Thomsen and Pedersen, 2005]. Assim, pode-se definir uma base de dados como uma coleção de dados, que são logicamente relacionados e que são geridos por um Sistema de Gestão de Base de Dados (SGBD) [PentahoCommunity, 2014]. Este SGBD é composto por um conjunto de programas que têm a missão de armazenar, gerir, organizar e proteger os dados. Para além disso, este sistema promove a partilha da informação nele contida a vários utilizadores e/ou aplicações, garantindo sempre a integridade dos dados [PentahoCommunity, 2014] [Hodak et al., 2007]. Por uma questão de simplicidade, muitas vezes é utilizado o termo base de dados para definir o conjunto: base de dados e SGBD. Contudo, quando se aborda os sistemas de bases de dados é importante referir que existem outros componentes para além do SGBD. Estes componentes podem dividir-se em três níveis diferentes [Bertino and Sandhu, 2005]:

- A nível de hardware - Existe um conjunto de dispositivos físicos responsável pelo armazenamento dos dados e funcionamento da base de dados (discos, processado- res, memórias, cabos de conexão, entre outros);
- A nível de software - Para além do conjunto de programas pertencentes ao SGBD, é comum os sistemas de bases de dados possuírem software de manipulação de dados (por exemplo, SQL (Structured Query Language) Developer e TOAD for Oracle);
- A nível de recursos humanos - São várias as pessoas envolvidas com os sistemas de base de dados (designer da base de dados, administrador do sistema, utilizadores).

A interligação destes componentes é um desafio de extrema importância para os designers e administradores, especialmente se o sistema for muito complexo [Tereso and Bernardino, 2011].

Nos anos 70 e com o objetivo de regular a implementação de sistemas de bases de dados e o relacionamento entre os seus componentes, surgiu a arquitetura ANSI-SPARC (American National Standards Institute - Standards Planning and Requirements Committee) [Bertino and Sandhu, 2005] [Chan, 2008]. Esta arquitetura é importante para demonstrar algumas funcionalidades do SGBD, nomeadamente na organização dos recursos. Esta arquitetura é, maioritariamente, denominada por arquitetura de 3 níveis (three-level architecture). Isto deve-se ao facto desta partir do princípio de que a estrutura de uma base de dados deve estar dividida em três camadas de abstração: a camada externa, a camada conceptual e a camada interna/física [Tereso and Bernardino, 2011] [Chan, 2008].

Monitorização e Prevenção de Falhas

A monitorização de falhas é um processo complexo, que é realizado através da avaliação de dados relativos ao comportamento do sistema. Essa avaliação é igualmente feita quando o sistema está exposto a fatores que possam originar uma falha, isto é, independentemente da área. O processo de monitorização de falhas consiste em obter informações úteis para evitar situações de futuras falhas [Shallahamer, 2007]. A prevenção de falhas está muitas vezes relacionada com o conceito de risco, assim, a sua importância é proporcional aos efeitos negativos que possam advir de uma possível falha [Shallahamer, 2007]. Esta característica torna a prevenção de falhas nas base de dados dos sistemas hospitalares uma questão essencial, pois está em jogo a qualidade dos serviços prestados aos cidadão.

Apesar da evolução dos mecanismos de tolerância a falhas, as falhas continuam a acontecer e podem ter várias origens, recursos físicos que deixem de funcionar e erros humanos na utilização do sistema são dois exemplos dessas falhas. Estas falhas conduzem, normalmente, a erros que afetam o comportamento normal do sistema. Todos os sistemas, quer sejam computacionais ou não, são caraterizados por um comportamento padrão que devem seguir, de forma a atingirem os objetivos para os quais foram construídos e implementados. Quando se verifica um comportamento distinto do esperado, diz-se que ocorreu um defeito no sistema [Yi, 2006]. Como já foi mencionado, um sistema de base de dados é composto por vários componentes, significando que um defeito no funcionamento do sistema pode advir de uma falha em qualquer um desses componentes [Yi, 2006].

Por vezes, estes defeitos originam a interrupção involuntária do sistema, a esta interrupção dá-se o nome de unplanned downtime [Hong, 2006]. No entanto, pode existir necessidade de interromper o funcionamento do sistema por outros motivos, por exemplo, manutenção ou atualização de componentes. Nestes casos, é agendada uma hora mais apropriada para que se possa parar o sistema com o menor nível de prejuízo. A este tipo de interrupções do sistema dá-se o nome de planned downtime [Godinho, 2011].

Aos sistemas que continuam disponíveis mesmo na presença de falhas em componentes de hardware (discos, memória, processador) ou software (sistema operativo, SGBD, aplicações) dá-se o nome de sistemas tolerantes a falhas [Sullivan, 1992].

1.2 Motivação e Objetivos

O Centro Hospitalar do Porto (CHP) possui um Data Center onde residem todos os seus Sistemas de Gestão de Bases de Dados. Só para se ter uma ordem de grandeza do volume de dados armazenados e nas transações realizadas, atente-se nos seguintes números:

- Número médio de sessões: 5696;
- Número acumulado de utilizadores: 27952;
- Número de sistemas de informação suportados: 34.

Dada a dinâmica associada a esta área de atuação e algum crescimento futuro (e.g., centro de ambulatório, nova maternidade) os valores apresentados irão seguramente sofrer um aumento.

Como se compreende, a disponibilidade, o desempenho e a confiança associados a estes sistemas são factores da maior importância num futuro próximo, trazendo desafios e oportunidades que estão na base desta proposta de trabalho.

Apesar dos avanços tecnológicos que se tem vivido nos últimos anos na área dos Sistemas de Gestão de Bases de Dados (SGBD), nos ambientes críticos com requisitos extremos ao nível da disponibilidade, segurança, confiança e desempenho como são os ambientes hospitalares, alguns aspectos ainda permanecem em aberto, proporcionando oportunidades de investigação [Abelha et al., 2004] [Sousa, 2002].

Sem exceção, todos os sistemas de informação integram bases de dados de dimensões consideráveis e contendo dados de vários tipos (desde valores numéricos até imagens em formato DICOM). Sistemas como o Processo Clínico Electrónico, o Sistema de Gestão Hospitalar e a Plataforma de Meios Complementares de Diagnóstica e Terapêutica e Interoperação de Sistemas, consomem a maior parte dos recursos de armazenamento, de processamento e de comunicação, não podendo estar sujeitos a falhas.

No plano estratégico dos sistemas de informação está prevista a criação de uma Data Warehouse para suporte a um sistema de Business Intelligence. Outros requisitos advêm da necessidade de se suportar processos de descoberta de conhecimento em bases de dados e data mining [Santos and Azevedo, 2005].

Ao nível da gestão dos SGBD são já utilizadas tecnologias para monitorização e afinação das configurações. A par disto, um conjunto de profissionais técnicos acompanha e tenta, de forma pró-activa garantir os níveis de serviço exigidos.

Os aspectos supra mencionados, combinados com a multiplicidade de tecnologias de SGBD instaladas, resultam num desafio verdadeiramente crítico, para o qual seria extremamente importante se fosse possível contar com modelos de previsão de eventos ou situações com alguma antecedência, e formas de atuação. A realização deste trabalho assume como principais objetivos:

- Prever a taxa futura de utilização de um determinado sistema e do volume de armazenamento necessário;
- Prever a ocorrência futura de um evento critico que implique a paragem de um SGBD e, consequentemente, todo o sistema de informação;
- Prever com antecedência os recursos necessários para solucionar um determinado problema.

1.3 Metodologias

O processo de investigação associado a este trabalho de doutoramento terá uma abordagem essencialmente quantitativa. As metodologias de investigação a utilizar serão de dois tipos: exploratória e confirmatória. A nível exploratório será efectuado uma abordagem geral que terá como principal objectivo a compreensão e caracterização do problema, a recolha de dados e a definição de modelos, através das metodologias baseadas em Estudo de Casos e Experiências de Campo. As técnicas a utilizar resumir-se-ão à aquisição de dados sobre o comportamento dos SGBD e da modelação de dados que contribuirá para a concepção do modelo de eventos e ações. Após a fase exploratória, seguir-se-á a fase confirmatória que tem como objectivo a prova de um conceito. Para isso serão utilizados os dados recolhidos e os modelos criados para através do desenvolvimento de um protótipo de sistema inteligente e de apoio à decisão poderem ser validados como artefactos, verificando qual o impacto das ações.

1.4 Publicações

Durante a realização deste projeto vários estudos foram realizados. Esses estudos foram submetidos a avaliação científica com sucesso, resultando nas seguintes publicações:

- Nuno Gonçalves, Cesar Quintas and José Machado,Clinical Business Intelligence to Prevent Stroke Accidents, in Applying Business Intelligence to Clinical and Healthcare Organizations, José Machado and António Abelha (eds), IGI Global, 2016.

- Filipe Portela, Alexandra Cabral, António Abelha, Maria Salazar, César Quintas, José Machado, José Neves and Manuel Filipe Santos, Knowledge Acquisition Process for Intelligent Decision Support in Critical Health Care. Healthcare Administration: Concepts, Methodologies, Tools, and Applications. IGI Global Book, 2014.

- Eva Silva, Ana Alpuim, Luciana Cardoso, Fernando Marins, Cesar Quintas, Filipe Portela, Manuel Santos, José Machado and António Abelha, Business Intelligence and Nosocomial Infection Decision Making, in Integration of Data Mining in Business Intelligence Systems, Ana Azevedo and Manuel Santos (eds), IGI Global Book, 2014.

- Cardoso, Luciana, Fernando Marins, César Quintas, Filipe Portela, Manuel Santos, António Abelha and José Machado. "Interoperability in Healthcare."Cloud Computing Applications for Quality Health Care Delivery. IGI Global, 2014. 78-101. Web. 23 Jul. 2014. doi:10.4018/978-1-4666-6118-9.ch005.

- Filipe Portela, Pedro Gago, Manuel Santos, José Machado, António Abelha, Álvaro Silva, Fernando Rua, César Quintas&Filipe Pinto. Implementing a Pervasive Real-time Intelligent System for Tracking Critical Events with Intensive Care Patients. IJHISI - International Journal of Healthcare Information Systems and Informatics, Volume 8, Issue 4. IGI Global. (2013) (SCIMago SJR 0.121-Q4).

- Filipe Portela, Alexandra Cabral, António Abelha, Maria Salazar, César Quintas, José Machado, José Neves and Manuel Filipe Santos, Knowledge Acquisition Process for Intelligent Decision Support in Critical Health Care. Information Systems and Technologies for Enhancing Health and Social Care. IGI Global Book, 2013.

- Filipe Portela, Alexandra Cabral, António Abelha, Maria Salazar, César Quintas, José Machado, José Neves and Manuel Filipe Santos, Knowledge Acquisition Process for Intelligent Decision Support in Critical Health Care. Information Systems and Technologies for Enhancing Health and Social Care. IGI Global Book, 2013.

- Filipe Portela, Pedro Gago, Manuel Filipe Santos, José Machado, António Abelha, Álvaro Silva, Fernando Rua, César Quintas, Filipe Pinto, Intelligent and Real Time Data Acquisition and Evaluation to Determine Critical Events in Intensive Medicine, Procedia Technology, Volume 5, 2012, Pages 716-724, ISSN 2212-0173.

- Paulo Silva, Cesar Quintas, Manuel Santos, António Abelha and José Machado, Step Towards Fault Forecasting in Hospital Information Systems, 5th IEEE International Conference on Biomedical Engineering and Informatics, Chongqing, China, 2012. DOI: 10.1109/BMEI.2012.6513041.

- Paulo Silva, Cesar Quintas, Pedro Gonçalves, Gabriel Pontes, Manuel Santos, António Abelha and José Machado, Intelligent Systems based in Hospital Database Malfunction Scenarios, 4ht IEEE International Conference on Industrial Engineering and Engineering Management, Hong Kong, 2012.

- Paulo Silva, Cesar Quintas, Júlio Duarte, Manuel Santos, José Neves, António Abelha and José Machado, Hospital Database Workload and Fualt Forecasting, 2012 IEEE EMBS Conference on Biomedical Engineering and Sciences, Langkawi, Malaysia, 2012. DOI: 10.1109/IECBES.2012.6498150.

- Cabral, A., Pina, C., Machado, H., Abelha, A. ,Salazar, M., Quintas, C., Portela, F, Machado, J., Neves, J., Santos, M., Data Acquisition Process for an Intelligent Decision Support in Gynecology and Obstetrics Emergency Triage, in Communications in Computer and Information Science, 221 CCIS (PART 3), Springer, 2011 (SCIMago SJR 0.14-Q4) DOI: 10.1007/978-3-642-24352-3_24.

- Pedro Gonçalves, Luciano Alves, Tiago Sá, Cesar Quintas, Miguel Miranda, António Abelha and José Machado, Object trajectory simulation - An evolutionary approach, in Novais, P.; Machado, J.; Rodrigues, César; Abelha, A.(ed), Modelling and Simulation 2011, EUROSIS.

- Júlio Duarte, Maria Salazar, Cesar Quintas, Manuel Santos, José Neves, António Abelha and José Machado, Data Quality Evaluation of Electronic Health Records in the Hospital Admission Process, 9th IEEE/ACIS International Conference on Computer and Information Science, ICIS 2010, Japão. DOI: 10.1109/ICIS.2010.97.

- Santos M, Portela F, Vilas-Boas M., Machado J., Abelha A., Neves J., Silva A., Rua F., Salazar M., Quintas C., Cabral A.F., Nursing Information Architecture for Situated Decision Support in Intensive Care Units, Recent Advances in Applied Informatics and Communications, WSEAS PRESS, 2009.

- Júlio Duarte, Miguel Miranda, António Abelha, José Machado, José Neves, Manuel Santos, Carlos Alberto, Maria Salazar, César Quintas, Alexandra Ferreira, João Neves, Agent based Group Decision Support in Medicine, in Proceedings of Worldcomp 2009, 2009 International Conference on Artificial Intelligence, Las Vegas, USA, 2009. ISBN: 978-160132109-1.

- Santos M, Portela F, Vilas-Boas M., Machado J., Abelha A., Neves J., Silva A., Rua F., Salazar M., Quintas C., Cabral A.F., Intelligent Decision Support in Intensive Care Units Nursing Information Modeling, Computers and

Simulation in Modern Science Volume III, WSEAS PRESS, 2009. ISBN: 978-960-474-256-1

Capítulo 2

Revisão da Literatura

Durante a realização deste trabalho realizaram-se várias pesquisas bibliogáficas. Estas pesquisas resultaram na revisão da literatura apresentada ao longo do atual capítulo. Para a realização destas pesquisas foram utilizadas as seguintes base de dados bibliográficas:

- *Isi Web of Knowledge*;
- *Science Direct*;
- *Google Scholar*;
- *PubMed*;
- *SpringerLink*;
- *Scopus*;

Para a realizaçaõ destas pesquisas bibliográficas foram utilizadas as palavras-chaves seguintes:

- Sistemas de Informação Hospitalar, Processo Clínico Eletrónico, Sistemas de Informação na Sade, Registo Clínico Eletrónico;
- Sistemas de Gestão de Bases de Dados, Manutenção de Bases de Dados, Esruturas de Base de Dados;
- Monitorização de Sistemas, Monitorização e Prevenção de Falhas, Falhas dos Sistemas;

Para além das pesquisas bibliográficas, o conhecimento adquirido com a experiência, proveniente do trabalho de gestão do Serviço de Informação do Centro Hospitalar do Porto, contribuiu também para realização deste trabalho, em especial para a percepção da realidade vivida num contexto regional e nacional. Igual contributo foi conseguido através de visitas a *sites* de instituições e organizações com responsabilidades sobre os temas supra citados.

2.1 Interoperailidade

No contexto das TI na saúde, a interoperabilidade é a capacidade de diferentes sistemas informáticos e aplicações de software comunicarem e trocarem dados com exatidão, eficácia e consistência, e de usarem a informação trocada [Kirsh, 2008] [Haux, 2006b].

Desta forma, a interoperabilidade tem por objectivo apoiar a [Duarte, 2015]:

- Transferência e partilha de dados em múltiplos locais ou larga escala empresarial;
- Transferência e integração de conhecimento;
- Transferência, mapeamento e integração de terminologia médica;
- Transferência de imagens;
- Integração com aplicações clínicas e não clínicas.

De forma a garantir a interoperabilidade foi necessário desenvolver normas de uniformização dos sistemas de PCE [Abelha, 2004] [Zaleski, 2009]. Sem interoperabilidade, os médicos, farmácias e hospitais não poderiam partilhar informação dos pacientes, o que é vital para efetuar cuidado remoto, centrado no paciente e em tempo útil. Existem atualmente muitos vendedores competitivos de sistemas de PCE, cada um vendendo uma suite de software que em muitos casos não é compatível com os dos seus adversários. Existem pelo menos 25 grandes marcas de software de PCE atualmente no mercado [Haux, 2006a].

2.1.1 Arquiteturas de Integração

Com vista à integração de sistemas heterogéneos, existem arquiteturas distintas a serem utilizadas, as quais com características que diante das suas disparidades, disponibilizam comportamentos diferentes. Entre as arquiteturas mais utilizadas devem ser referenciadas [Wei, 2012]:

- Arquitetura Ponto a Ponto;
- Arquitetura Hub and Spoke;
- Arquitetura Distribuída;
- Arquitetura Orientada a Serviços;

Arquitetura Ponto a Ponto

As ligações ponto a ponto, com o intuito de estabelecer circuitos de comunicação, integrando numa rede de aplicações diferentes soluções independentemente das empresas envolvidas, são passíveis de ser implementadas a curto prazo de forma bastante eficiente [O'Brien, 2009]. Baseiam-se num canal, metodologia, sintaxe e semântica própria que liga dois pontos (aplicações) de um sistema.

Devido às descritas ligações forte ponto a ponto, a sua forte capacidade de optimização e a eficiência inerente a estas restrições, continuam a ser bastantes utilizadas nos dias de hoje e muito provavelmente continuarão a ser a curto-prazo. No entanto, a eficiência derivada das interligações fortes, estáticas e muitas das vezes proprietárias entre os sistemas envolvidos, é incomportável num sistema de grande dimensão.

Noutras palavras, se bem que as ligações individualmente sejam simples e pouco complexas, uma das principais desvantagens das ligações ponto a ponto, é que à medida que o número de aplicações envolvidas aumenta, também aumenta a complexidade global do ambiente no qual se inserem, algo muito visível na Figura 2.1. Tendo em consideração a heterogeneidade e a variedade de soluções e serviços que fazem parte de uma unidade prestadora de cuidados de saúde, a adopção desta

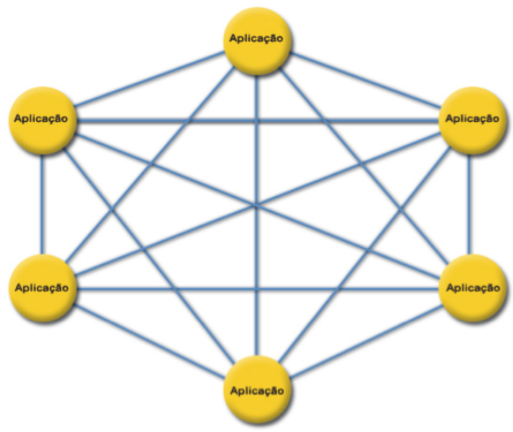

Figura 2.1: Arquitetura Ponto a Ponto (adaptado de [Aier and Schönherr, 2006])

arquitetura pode resultar em custos altos de manutenção, falta de flexibilidade na alteração, melhoria e substituição deste pontos e na adição de novos pontos nesta rede [O'Brien, 2009].

Arquitetura *Hub and Spoke*

O nome desta arquitectura advém da sua representação, na qual uma entidade central (*hub*) se liga através dos seus raios (*spokes*) a diferentes aplicações (Figura 2.2). As primeiras tecnologias formais de integração funcionavam segundo o princípio que toda a informação proveniente das diferentes aplicações no sistema teria que ser processada por uma unidade central, um *Hub* no qual se concentrava toda a informação. Esta unidade central processa todas as comunicações entre elementos do sistema, gerindo todo o fluxo e conteúdo das mensagens trocadas, mapeando os elementos do sistema e os caminhos de comunicações [O'Brien, 2009].

Ao contrário das ligações ponto a ponto, as aplicações ligadas ao *Hub* podem ser alteradas de forma independente umas das outras. Isto é, como não comunicam entre si, somente com a unidade central, se na unidade central um novo ponto ocupar o "lugar" de outro, esta alteração será invisível para os todos os outros pontos que

Figura 2.2: Arquitectura Hub (adaptado de [Aier and Schönherr, 2006])

se encontram ligados ao *Hub*. Este processo dá a liberdade de escolha e a melhoria das aplicações presentes no sistema, ao mesmo tempo que diminui a complexidade do ambiente de integração comparativamente com as ligações ponto a ponto quando o número de ligações aumenta.

No entanto, a centralização de processamento torna mais complexa a criação inicial de comunicação via *Hub*, havendo a necessidade das aplicações terem conhecimento umas das outras para existir comunicação, algo que pode muito bem dificultar o processo de adicionar e remover pontos do sistema. Ademais, a centralização das comunicações exige um enorme poder de processamento e armazenamento, sendo também um ponto de falha do sistema. Isto é, a carga sobre esta unidade central exige altas características para este equipamento, assim como torna quase impossível quaisquer melhorias profundas, pois este processo causaria a quebra de todo o sistema. Pontos centrais redundantes e com balanceamento podem ajudar a solucionar este problema, no entanto, os problemas centrais a nível de modularidade e escalabilidade mantém-se.

Arquitetura Distribuída

Tendo em atenção os problemas de escalabilidade da arquitectura *Hub and Spoke*, estes podem ser abordados de outra forma, realizando a troca, monitorização, divisão, combinação e "tradução" de mensagens em camadas individuais de abstracção e mais próximas da aplicação em termos lógicos. Estas camadas distintas de *middleware* encontram-se ligadas á aplicação da qual são responsáveis e processam, traduzem, enviam e recebem as comunicações com outras aplicações. O resultado desta arquitectura é uma distribuição das aplicações e *middlewares*, descentralizando o processamento das diferentes aplicações da unidade prestadora de cuidados de saúde (Figura 2.3).

Figura 2.3: Arquitetura Distribuída (adaptado de [Aier and Schönherr, 2006])

A distribuição e processamento nestas unidades são governados por um conjunto de regras centralizadas e pelos requisitos do fluir do trabalho no sistema. A maior parte do processamento relativo à integração é localizado no *middleware*, permitindo uma distribuição acessível do trabalho no sistema, assim como o crescimento dinâmico do sistema sem o elevado número de pontos envolvidos aumentar a complexidade desta tarefa. Esta capacidade representa uma considerável melhoria relativamente às arquitecturas de Ponto a Ponto e *Hub and Spoke* [O'Brien, 2009].

Inicialmente, existia um conjunto de tecnologias distribuídas que não interoperavam entre si, exigindo uma homogeneidade nos sistemas distribuídos implementadas. Tecnologias como o CORBA, o COM da *Microsoft* e o *Java* RMI regiam as suas próprias metodologias de sistemas distribuídos havendo, no entanto, algumas tentativas de implementação de ligações entre estas [O'Brien, 2009].

Arquitetura Orientada a Serviços (SOA)

A ideia central da arquitectura orientada a serviços tem como princípio chave a resolução de problemas, a divisão destes em unidades mais simples. No entanto, ao contrário da metodologia comum a grande parte dos algoritmos de resolução de problemas complexos, nesta arquitectura procura-se a independência ente estas unidades que em conjunto resolvem um problema. Por outras palavras, é procurado a divisão segundo áreas de automação lógica de forma a unitariamente disponibilizarem um serviço parte de um outro serviço mais complexo, mas em que qualquer destas unidades pode ser ocupada por outra com as mesmas funcionalidades. Estas unidades base podem ser distribuídas, providenciando um serviço de forma distribuída e dissociada [Erl, 2004]. Se bem que à primeira vista possa se assemelhar à arquitectura distribuída, a diferença está na dependência entre diferentes unidades distribuídas, algo comum em arquitecturas distribuídas, mas contra o ideal de uma arquitectura verdadeiramente orientada a serviço.

Através do paradigma SOA um sistema não estará dependente das suas unidades, isto é, serviços poderão ser alteradas facilmente e ser substituídos por outros que disponibilizem o mesmo serviço, permitindo modularidade, escalabilidade e independência para quem utiliza esta arquitectura [Juric et al., 2007].

2.1.2 *Standards* de Comunicação

Para a compreensão da informação trocada entre diferentes entidades, sejam estas humanas ou virtuais, é necessário o conhecimento por parte das duas partes da sintaxe e da semântica da mensagem trocada. Desta forma, a nível de cada área podem surgir diferentes sintaxes e semânticas para a troca de mensagens. Por outras palavras, não só a estrutura das mensagens e o tipo de campos nestas contidas são necessários para a completa compreensão de uma mensagem. Isto é, para a inexistência de ambiguidade, também os significados, o contexto e as relações entre os diferentes termos, têm de ser conhecidos e utilizados por ambas as partes na comunicação. Na área da saúde e seguindo este raciocínio de divisão de standards de comunicação em dois níveis, podem ser referenciados múltiplos projetos e subsequentes normas internacionais.

Health Level Seven **(HL7)**

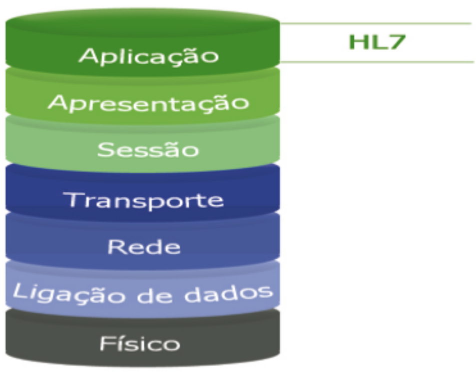

Figura 2.4: Protocolo OSI (adaptado de [Dolin et al., 2001])

O HL7 é um protocolo de comunicação na camada sete da pilha OSI (Figura 2.4), relativo ao nível de estruturação de mensagens contendo informação médica. Este protocolo encontra-se atualmente em rápida disseminação a nível de aplicações baseadas em software e quase que horizontalmente utilizado por todas a áreas de instrumentação médica. A informação acerca desta norma é bastante restrita, não havendo muitas fontes de informação e de avaliação [Smith and Ceusters, 2007]. No entanto, encontra-se embebida em muitos equipamentos médicos, sendo essencial para outras aplicações baseadas em software receberem e analisarem de forma conveniente a sua informação [Turham et al., 2008].

Além do protocolo HL7 a *framework* HL7-RIM providencia uma representação das estruturas e das relações entre unidades de informação independentes da fonte, das tecnologias específicas ou do ambiente de implementação. é desta forma desenhado para suportar o trabalho de desenvolvimento no esquema de uma base de dados e em todo o processo de engenharia de *software* através do estabelecimento de um único ambiente de comunicação que pode ser partilhado entre qualquer sistema de informação da instituição e os subsequentes serviços [Dolin et al., 2001].

Por conseguinte, o protocolo HL7 procura providenciar linhas orientadoras para a interoperabilidade entre sistemas distintos de forma a melhorar a qualidade dos cuidados de saúde, reduzir a ambiguidade na informação e melhorar a qualidade da transferência de conhecimento entre os utilizadores deste protocolo. Relativamente a diferentes versões do protocolo, uma versão muito utilizada continua a ser a 2.4, havendo no entanto uma boa utilização e melhorias na versão 3 deste protocolo [Sujansky et al., 2009]. O objectivo central da versão 3 encontra-se fortemente ligada ao *Reference Information Model* (RIM), responsável por fornecer um conjunto de linhas orientadoras que são bem definidas, testáveis, e de permitir certificar a conformidade dos vendedores. Esta nova versão utiliza uma metodologia de desenvolvimento orientada a objetos e um RIM para criar mensagens. O RIM permite uma representação explicita das ligações semânticas e lexicais, que existem entre a diferente informação contida nos campos da mensagem seguindo esta versão 3 do protocolo HL7 [Turham et al., 2008].

Através deste protocolo, variadas tarefas associadas a equipamentos médicos, desde um agendamento num RIS ao envio de resultados de um leitor de RFID monitorizando o movimento de um paciente, conseguem enviar e receber informação de qualquer SI numa instituição de saúde, podendo integrar a informação gerada compor estes equipamentos sem a utilização de desenvolvimento de *middlewares* proprietários para cada um destes.

Esta iniciativa é muito semelhante à desenvolvida por grupos de investigação e implementação de sistemas multi-agente através do estabelecimento de protocolos e standards para comunicação entre diferentes sistemas baseados em agentes. Como exemplo, podemos considerar as normas desenvolvidas pela *Foundation for Intelligent Physical Agents*, de forma mais concreta, a sua *Agent Comunication Language*, que será abordada na seguinte secção.

2.2 Plataforma Integradora AIDA

A plataforma AIDA [Abelha, 2004] foi desenvolvida de forma a apoiar a integração e difusão da informação gerada em ambiente hospitalar, por todos os serviços existentes numa instituição prestadora de cuidados de saúde. Esta plataforma incorpora diferentes paradigmas de integração, utilizando princípios SOA e de sistema multi-agente de forma a integrar a informação dos diferentes parceiros de negócio da instituição. Desta forma é mantida a independência e a modularidade do SOA e à inteligência e autonomia associada á inteligência artificial [Abelha et al., 2004].

Em face ao que foi descrito anteriormente relativamente ao que define um sistema de informação da saúde, a AIDA ocupa uma posição de relevo na implementação e no funcionamento de um SIS. A capacidade de permitir a comunicação de diferentes serviços e a não centralização destes é vital para evitar a utilização de ligações ponto a ponto que restringem o crescimento da infra-estrutura associada ao SIS [Duarte, 2008]. A não modularidade de serviços torna as alterações no sistema muito complexas, aumentando os custos globais do sistema de informação é por conseguinte compreensível a atual preocupação dos órgãos e entidades responsáveis pelo financiamento e pela regulação destes projetos associados à compra ou desenvolvimento de novos SIS, com questões de flexibilidade na interoperabilidade e integração de sistemas heterogéneos [Berg, 2009][Machado et al., 2010].

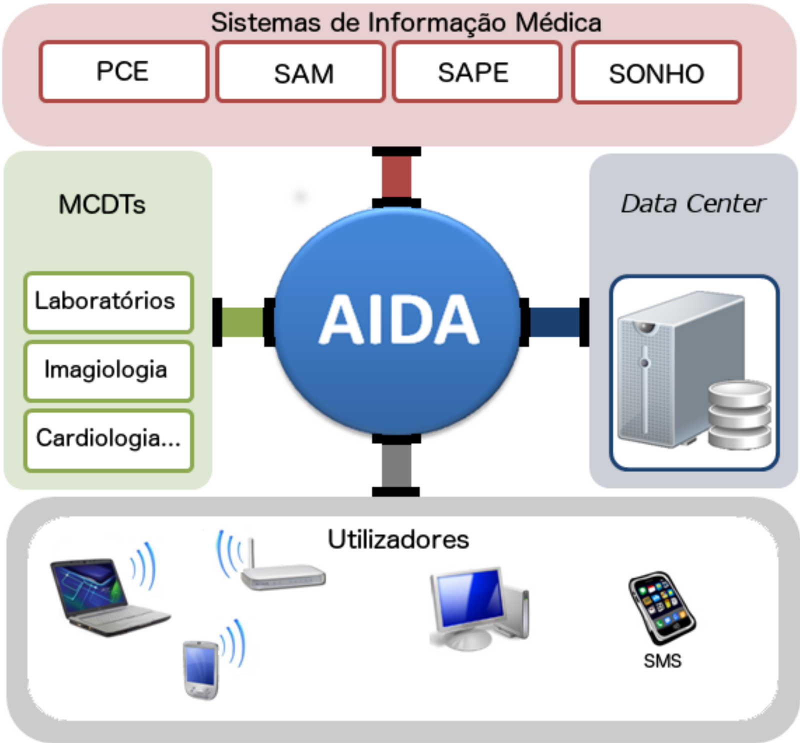

Figura 2.5: Arquitetura da Plataforma AIDA. [Duarte, 2008]

A plataforma AIDA, permite uma integração dos diferentes sistemas de informação clínicos e administrativos existentes no hospital e futuros, numa abordagem global, evitando-se a ligação ponto a ponto. é possível desta forma interligar as aplicações existentes e ainda potenciar a interconexão das unidades hospitalares, centros de saúde e outros organismos prestadores de cuidados de saúde.

A arquitetura da plataforma AIDA é apresentada na Figura 2.5, de uma forma sintetizada. Ao nível de uma Unidade Hospitalar permite [Abelha et al., 2007] [Abelha et al., 2002]:

- Arquivar em formato PDF e XML os meios complementares de diagnóstico como análises laboratoriais, imagens médicas (e.g., TAC, RX, ECO) e outros exames médicos (e.g., cardiológicos, função respiratória, broncoscopia);

- Manter uma base de dados de ocorrências e erros;
- Manter uma base de dados de *logs* como suporte à auditoria tendo como objectivo a possibilidade de conhecer com exatidão a informação consultada pela varias categorias de utilizadores que acedem á plataforma;
- Difundir pela intranet do hospital através da utilização de um *Internet Browser* (por exemplo, *Netscape*, *Internet Explorer*) ou através de outros sistemas os meios complementares de diagnóstico existentes;
- Implementar um esquema de segurança ao nível dos utilizadores e perfis de utilização (inclui o recurso à tecnologia biométrica);
- Manter um conjunto de tabelas de conversão entre os vários códigos utilizados nos sistemas a interligar (e.g., tabela de faturação do HIS);
- Implementar um conjunto de funcionalidades que são transversais aos vários sistemas como:
  - Disponibilização de dados para o módulo de faturação do HIS;
  - Requisição de meios complementares de diagnóstico e gestão de pendentes;
  - Disponibilização da informação dos doentes contida na base de dados do HIS para os outros sistemas;
  - Registo e upload de relatórios de exames através de um interface *WEB*;
  - Integrar os vários sistemas como:
    - Aplicações de gestão laboratorial
    - Aplicações de aquisição e tratamento de imagem médica;
    - Aplicações de aquisição e tratamentos de dados anestésicos;
    - Aplicações de registo de exames médicos;
    - Portais cooperativo;

    * Prescrição electrónica;
    * *HelpDesk*;

- Disponibilizar um conjunto parametrizável de estatísticas e índices de produtividade;

- A passagem futura de informação sobre meios complementares de diagnóstico e clínica entre as unidades hospitalares, os centros de saúde e os organismos privados (de acordo com regras a definir).

No desenvolvimento da plataforma AIDA estiveram presentes requisitos como a escalabilidade, a minimização dos custos de funcionamento, manutenção e armazenamento. Assim como foram tomados em conta aspectos que têm a ver com a simplicidade e independência dos mecanismos e dos protocolos de integração utilizados.

2.3 Processo Clínico Eletrónico

O sistema associado ao Registo Clínico Electrónico Orientado ao Paciente (RCEOP) tem como objectivo o colmatar das necessidades de sistematização, registo e análise de dados associados à prática e investigação clínica. Trata-se de um registo rigoroso de dados do doente/utente, que possibilita a prestação de cuidados continuados, a comunicação entre membros da equipa, permitindo que em qualquer momento qualquer outro profissional de saúde assuma a responsabilidade pelos cuidados do doente/utente, evite erros de identificação e facilite a obtenção de dados para investigação, formação e auditoria [Duarte et al., 2011].

A quantidade e a qualidade da informação disponível para os profissionais de saúde nos cuidados de saúde tem impacto tanto no resultado como na continuação da prestação dos cuidados de saúde. A informação incluída nos RCEOP tem diferentes funções no processo de tomada de decisão no cuidado ao utente, e também suporta o processo de tomada de decisão na gestão e na política de saúde [Duarte, 2008].

Um dos principais objectivos deste projeto tem sido a substituição de documentação armazenada em papel por documentos electrónicos, aumentando a quantidade de dados processados e reduzindo os custos e tempo gasto neste processo, de forma a tornar a assistência ao paciente mais eficiente e melhorar a qualidade do serviço prestado [Machado et al., 2006]. O conhecimento obtido é também utilizado para propósitos educacionais e de treino, assim como para alimentar sistemas inteligentes ubíquos para apoio a médicos, enfermeiros e técnicos [Duarte, 2008].

Os sistemas de registo clínico electrónico foram classificados com base na definição da *International Organization for Standardization* (ISO). A ISO também define um número de termos habitualmente utilizados para descrever diferentes tipos de software de registo clínico electrónico [Lippeveld et al., 2000] [ISO, 2005]:

- Registo Médico Eletónico (EMR, do inglês *Electronic Medical Record*) - principalmente focado nos cuidados de saúde;
- EMR Departamental - contém informação introduzida por um departamento hospitalar único, *Picture Archiving and Communication System* (PACS), registos de anestesia, cuidados intensivos, ambulatório, emergência, cardiologia e muitos outros.
- EMR Inter-departamental - contêm informação sobre dois ou mais departamentos hospitalares;
- EMR Hospitalar - contém toda ou a maior parte da informação clínica do utente num hospital em particular;
- EMR Inter-hospitalar - contém a informação clínica do utente de dois ou mais hospitais;

- Registo Electrónico do Utente (EPR, do inglês *Electronic Patient Record*) - contém toda ou a maior parte da informação clínica do utente num hospital em particular;
- Registo Computorizado do Utente (CPR, do inglês *Computerized Patient Record*) - contém toda ou a maior parte da informação clínica do utente num hospital em particular;
- Registo de Cuidados de Saúde do Utente (EHCR, do inglês *Electronic Health Care Record*) - contém toda a informação sobre a saúde do utente;
- Registo Pessoal de Saúde (PHR, do inglês *Personal Health Record*) - controlado pelo paciente, contém informação introduzida em parte por ele próprio;
- Registo Médico Computorizado (CMR, do inglês *Computerized Medical Record*) - criado por digitalização de um registo de saúde baseado em papel;
- Registo Médico Digital (DMR, do inglês *Digital Medical Record*) - um registo baseado na web com informação mantida pelo prestador de cuidados de saúde;
- Repositório de Dados Clínicos (CDR, do inglês *Clinical Data Repository*) - um serviço de alojamento de dados operacional que aloja e gere os dados clínicos obtidos a partir de diferentes prestadores de cuidados de saúde.
- Registo de Electrónico do Cliente (ECR, do inglês *Electronic Client Record*) - o âmbito é definido por profissionais de cuidados de saúde que não médicos, tais como fisioterapeutas.
- EMR Virtual - sem definição autoritativa.
- Registo de Saúde da População (PHR, do inglês *Population Health Record*) - contém dados agregados e usualmente não-identificada.

Na metodologia de registo nestes sistemas podem ser denotados diferentes paradigmas que dão diferentes perspectivas sobre a mesma informação. Nos registos médicos electrónicos orientados ao tempo, os dados são apresentados de forma cronológica. Quando orientados ao problema, são inseridas notas para cada problema atribuído ao utente, e cada problema é descrito de acordo com informação Subjetiva, Objectiva, Avaliativa e Planeada (SOAP). No registo orientado à fonte, o conteúdo do registo é arranjado de acordo com os métodos através dos quais a informação foi obtida como, por exemplo, notas de visitas, relatórios raios-X e análises ao sangue. Dentro de cada secção, os dados são reportados de forma cronológica [Tange et al., 2001] [Haux, 2006a] [Kalra, 2006] [Weed, 2008] [Dick and Steen, 1991].

O RCEOP irá funcionar em intranet, na rede interna do CHP, com protocolos de acesso e não num sistema de partilha de dados. Nele é necessário inserir mecanismos de segurança adequados e acessos condicionados a definir, sendo a informação confidencial encriptada como metodologia de segurança. Note-se que a implementação deste módulo no projeto implica uma forte componente na área da formação.

2.3.1 Requisitos do PCE

Atualmente existe um vasto leque de soluções ao nível do processo clínico electrónico, no entanto, a escolha de uma solução deve estar de acordo com os objectivos pretendidos e face aos sistemas de informação existentes. Algumas características podem ser destacadas [Weed, 2008] [Duarte, 2015] [Duarte et al., 2011]:

- Capacidade de interoperação com os sistemas de informação existentes e futuros;
- Adaptado ao panorama de prestação de cuidados de saúde em Portugal;
- Baseado em tecnologias web, não exigindo alterações nos postos de trabalho;
- Disponibilidade permanente na rede / ubiquidade;

- Utilização de tecnologias extensíveis e adaptáveis às particularidades dos profissionais de saúde;
- Suporte eficiente e direcionado ao panorama nacional a auditorias e processos legais;
- Suporte explícito à gestão, auditoria e avaliação;
- Rapidez e segurança no acesso à informação consolidada do doente e terapia, meios complementares de diagnóstico e prescrição;
- Promover o aumento da qualidade do serviço de saúde prestado;
- Permitir a diminuição da redundância de procedimentos, quer administrativos quer clínicos;
- Suporte a ações de prevenção e promoção da saúde;
- Articulação do trabalho dos profissionais de saúde e suporte à continuidade dos cuidados por estes prestados.

2.4 Sistemas de Base de Dados

O Hospital Geral de Santo António tinha, em 2004, um ambiente heterogéneo com ilhas de base de dados criadas e disseminadas pelos diferentes serviços clínicos e de apoio.

A partir desse ano, num primeira fase, foi efectuado um processo de consolidação das base de dados nos ambientes Microsoft SqlServer e Oracle. A principal preocupação foi garantir a segurança e disponibilidade dos dados sem grandes mudanças ao nível do código das aplicações. Num segundo momento, foi efectuado uma consolidação aplicacional com recurso a ferramentas para conversão do código aplicacional, normalmente Microsoft Access e DBase, em ambiente *Web PHP2.0*, *Web*

PHP3.0 e *Oracle Application Express* (Oracle APEX). Algumas destas conversões só foram possíveis com a reprogramação quase total da aplicação.

Com o desenvolvimento e implementação do Processo Clínico Electrónico (PCE), a maior parte dos sistemas legados foram integrados no PCE, e foi desenvolvidas ferramentas especificas para converter os relatórios de Alta, Admissão e Meio Complementar de Diagnóstico das aplicações legadas em formato PDF e XML e posteriormente, integrado no PCE com o numero de utente devidamente referenciado. Com estes passos, os sistemas legados foram totalmente desativados. Com a informatização total do HGSA e posteriormente com a criação do Centro Hospitalar do Porto, EPE (CHP), os sistemas foram ficando mais complexos e abrangentes, com a necessidade premente de interoperação de sistemas. A suportar toda a camada aplicacional do CHP, temos um conjunto de base de dados algumas consolidadas, outras em nodos isolados para as base de dados de grande dimensão.

Apresenta-se na Tabela 2.1 uma caracterização das base de dados associadas aos vários sistemas aplicacionais.

Tabela 2.1: Base de Dados do CHP

Consolidado	**Aplicação**	**Sessões**	**Utilizadores**
Sim	Sistema de Gestão de Doentes Hospitalares	500	1000
Sim	Sistema de Apoio ao Médico	500	1200
Sim	Sistema de Apoio à Prática de Enfermagem	300	1400
Sim	Processo Clínico Eletrónico	300	2000

Consolidado	**Aplicação**	**Sessões**	**Utilizadores**
Sim	Meios Complementares de Diagnósticos	300	70
Sim	Sistema de Informação de Radiologia	100	100
Sim	Sistema de Informação Laboratorial	10	10
Sim	Portal Interno	400	2500
Sim	Registo dos Electrocardiogramas	20	50
Sim	Analises de Anatomia Patológica	20	40
Sim	Monitorização	30	10
Sim	Registo das Sessões de Formação	50	4500
Não	Logística/Farmácia/Circuito de medicamento	400	1600
Não	Registo e validação dos resultados das analises	50	60
Não	Requisição e Agendamento Análises	200	1400
Não	Gestão Documental	50	70
Não	Diretório do Domínio de Rede	2000	5000
Não	*Holters*	3	5

Consolidado	**Aplicação**	**Sessões**	**Utilizadores**
Não	Avaliação de Desempenho	3	5
Sim	Endo e Vídeo Endoscopias	5	50
Sim	Uroscopias	1	10
Sim	Colposcopias	1	5
Sim	Broncoscopia-pneumoscopia	1	3
Sim	Gestão de sangue	50	150
Sim	Sistema de Gestão de Filas de Espera	30	10
Sim	Gestão de Assiduidade	100	5000
Sim	Hemodinâmica	2	4
Sim	Gestão de Obstetrícia e Ginecologia	5	15
Sim	Gestão de Visitas	2	15
Não	Sistema de Gestão de Recursos Humanos	50	60
Não	Gestão de Autenticação	100	800

Consolidado	**Aplicação**	**Sessões**	**Utilizadores**
Não	Registo de Informação clinica Urgência	100	800
Não	*Data Warehouse*	3	5
Não	OBIEE	10	5

2.5 Sistemas de Suporte à Decisão

Os Sistemas de Suporte à Decisão (DSS) têm por base modelos de tomada de decisão que analisam um elevado conjunto de variáveis que permitam a resposta a uma determinada questão. Ao longo dos anos estes sistemas têm vindo a evoluir, bem como a caracterização dos mesmos. De entre as várias descrições salientamos que para além de ser "um sistema computacional que auxilia o processo de tomada de decisão" [Finlay, 2004], é também "um sistema de informação interativo, flexível e adaptável, especialmente desenvolvido para o apoio à solução de um problema não estruturado e para o aperfeiçoamento da tomada de decisão [Turban, 2004]. Utiliza dados, tem uma interface amigável e permite ao utilizador ter a sua própria percepção na tomada de decisão". Assim e tendo por base estas descrições, podemos definir DSS como um sistema que dá a possibilidade aos seus utilizadores de conseguirem um rápido acesso à informação e permite a realização de análises, configurando-a às necessidades de cada um e de cada problema [Shohoud, 2002] [Rigor et al., 2008] [Machado et al., 2006].

2.5.1 Características mais comuns dos DSS

Podemos então verificar algumas das características essenciais dos sistemas de suporte à decisão [Marakas, 1998]. Estas características são partilhadas quer pelos sistemas de BI quer pelos sistemas *Adaptive Business Intelligence*:

- Sistemas aplicados em contextos de decisão semi-estruturados ou desestruturados;
- Direcionados para apoiar os gestores e não substitui-los;
- Apoiam todas as fases do processo de decisão;
- Estão focalizados para a eficácia da decisão e não para a eficiência;
- Estão sob controlo dos utilizadores do DSS;
- Têm subjacentes modelos e informação;
- Facilitam a aprendizagem por parte dos decisores;
- São interativos e de fácil utilização;
- Em regra são desenvolvidos através de um processo interactivo e com possibilidades de evolução;
- Providenciam apoio a todos os níveis de gestão, desde gestores de topo a gestores operacionais;
- Podem suportar a decisão múltipla independente ou interdependente;

Apoiam a decisão individual, de grupo, ou contextos de tomada de decisão em equipa.

2.5.2 Apoio à Decisão em Medicina

A disseminação das tecnologias de informação a nível da saúde tem sido crescentemente visível nas práticas diárias relacionadas com o registo clínico do paciente, havendo também resultado em soluções de apoio às decisões a estas associadas [Miranda et al., 2008][Dupuits and Hasman, 1995][Dreiseitl and Binder, 2005]. Os benefícios espectáveis associados a estas tecnologias de informação, encontram-se relacionados com [Dreiseitl and Binder, 2005]:

- A melhoria da qualidade do tratamento providenciando alertas automáticos e verificações da consistência da informação;
- Um aumento da eficácia por providenciar melhores normas de prática clínica e de registo; um aumento do conhecimento disponível com o providenciar da informação pretendida, quando e onde necessário [WHO, 2014]; e
- Redução de custos a longo prazo na gestão da informação e realização de meios complementares de diagnóstico, eliminando a necessidade de exames redundantes.

Com a implementação de diferentes normas ao nível de comunicação e armazenamento de dados, como HL7, SNOMED-CT, UMLS, têm surgido consideráveis avanços nesta área, faltando no entanto alguma compreensão da melhor forma de interagir com o utilizador e ultrapassar as barreiras iniciais relacionadas com a utilização de novas ferramentas. Surge desta forma a necessidade de avaliar continuamente estas características a principio mais subjetivas, assim como a influência destes sistemas de apoio a decisão nos intervenientes na prestação de cuidados de saúde [Duarte, 2015]. Estudos atuais indicam que uma pequena percentagem dos utilizadores, perante uma decisão, quando apresentados com uma recomendação de um sistema de apoio à decisão contrária à sua primeira decisão, alteram esta posição em face a justificação [Dreiseitl and Binder, 2005].

No entanto, as questões ético-morais associadas a estes sistemas e a problemática da responsabilização do erro médico, podem ainda limitar a disseminação destas tecnologias. O seu potencial justifica, contudo, uma aposta na sua investigação e no seu desenvolvimento.

2.5.3 *Business Intelligence*

Pode-se chamar *Business Intelligence* (BI) a tudo que é tecnologias, aplicações ou práticas que armazenam, integram, analisam e apresentam a informação do negócio. O BI é utilizado como suporte à tomada de decisão, dando a possibilidade de o utilizador tomar a melhor decisão. BI representa-se por um conjunto de conceitos e métodos que melhoram as capacidades dos DSS (sistemas de suporte à decisão). Os dados são armazenados diariamente em forma de bits, números, símbolos e objetos. A informação tem por base os conceitos, aprendidos e percebidos. Na Figura 2.6 podemos verificar a forma como negócio, a gestão e as tecnologias de informação se relacionam, salientando para o facto da existência de um conceito comum aos três, o BI. Os sistemas BI utilizam a informação que se encontra armazenada em bases de dados, e que, como demonstra a Figura 2.6, é tratada (ETP) e analisada, devolvendo no final os relatórios que dão suporte à tomada de decisão.

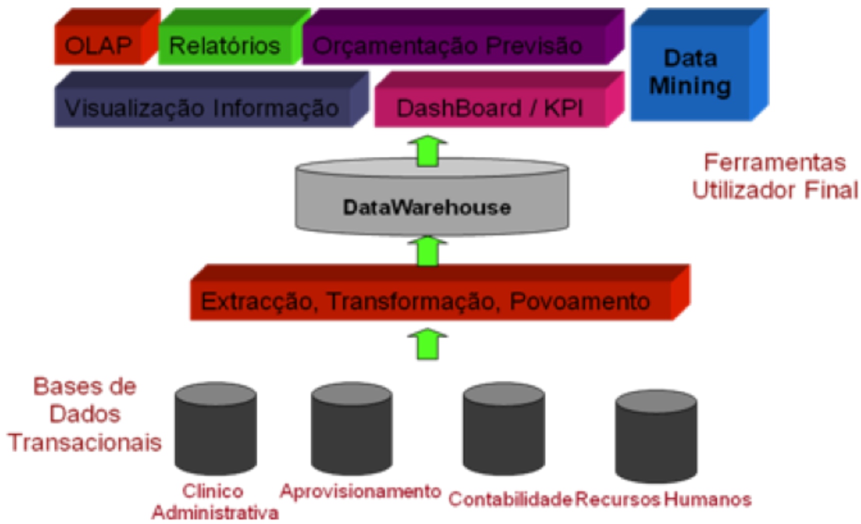

Figura 2.6: Relao entre negcio, gesto, tecnologias e BI

No entanto, e apesar dos benefícios que se obtinham o facto é que nem com todo o conhecimento do mundo se garantia a melhor decisão [30]. A existência de conhecimento apenas aumentava a nossa confiança, não melhorava a precisão das nossas decisões. Além disto a falta de processos inteligentes era clara, ou seja, era necessário que alguns dos processos/atividades fossem actualizáveis de uma forma automática, sem a intervenção humana, e adaptáveis às constantes mudanças. Na seguinte imagem podem ser verificadas quais as soluções/sistemas que suportam/utilização conhecimentos BI.

2.5.4 MEWS

O *Modified Early Warning Score* (MEWS) é um score simples e fisiológico, que pode permitir a melhoria na qualidade e segurança da gestão fornecida aos doentes de cirurgia [Gardner-Thorpe et al., 2006], sendo adequada a sua aplicação à cabeceira do doente [Subbe et al., 2001]. O principal objectivo é evitar atraso nas intervenções ou na transferência de doentes em estado crítico [Gardner-Thorpe et al., 2006], identificando os doentes que se encontram numa situação mais complicada [Subbe et al., 2001]. Para esse efeito é utilizada uma escala de pontuações que é calculada em função dos resultados obtidos de quatro parâmetros fisiológicos (pressão arterial sistólica, frequência cardíaca, frequência respiratória e temperatura corporal) e dos níveis de consciência (AVPU) (Alerta, Voz, Dor, Sem resposta). O MEWS utiliza uma escala de 0 (intervalo normal) a 3 (risco extremo) para classificar cada um dos resultados obtidos. A Tabela 2.2 demonstra essa classificação:

Tabela 2.2: MEWS

Score	**3**	**2**	**1**	**0**	**1**	**2**	**3**
PA Sistólica	$\leq$70	71-80	81-100	101-199	-	$\geq$ 200	-
FC (bpm)	-	$\leq$40	41-50	51-100	101-110	111-129	$\geq$130
FR (rpm)	-	$\leq$8	-	9-14	15-20	21-29	$\geq$30
Temp (°C)	-	$\leq$34.9	-	35-28.4	-	$\geq$38.5	-
AVPU	-	-	-	A	V	P	U

A soma total de cada parâmetro permite aferir o estado de gravidade de um doente [0-14], sendo que o nível mínimo de gravidade varia consoante o local onde é aplicado. Na UCI do CHP, segundo o Doutor Álvaro Silva, os doentes com um MEWS total igual ou superior a 8 são considerados de alto risco.

Capítulo 3

Resultados

3.1 I - Hospital database workload and fault forecasting

Paulo Silva, César Quintas, Júlio Duarte, Manuel Santos, José Neves, António Abelha and José Machado

2012 IEEE EMBS Conference on Biomedical Engineering and Sciences, Langkawi, Malaysia, 2012.

3.1.1 Abstract

With the growing importance of hospital information systems, databases became indispensable tools for day-to-day tasks in healthcare units. They store important and confidential information about patients clinical status and about the other hospital services. Thus, they must be permanently available, reliable and at high performance. In many healthcare units, fault tolerant systems are used. They ensure the availability, reliability and disaster recovery of data. However, these mechanisms do not allow the prediction or prevention of faults. In this context, it emerges the necessity of developing a fault forecasting system. The objectives of this paper are monitoring database performance to verify the normal workload for the main database of Centro Hospitalar do Porto and adapt a forecasting model used in medicine into the database context. Based on percentiles it was created a scale to represent the severity of situations. It was observe that the critical workload period is the period between 10:00 am and 12:00 am. Moreover, abnormal situations were detected and it was possible to send alerts and to request assistance.

3.1.2 Introduction

The evolution of technology contributed to the exponential growth of data which needs to be stored for the proper functioning in organizations. Databases are powerful tools where this large amount of data can be stored and managed in a simple way. The database management system (DBMS) coordinates a set of hardware and software components, furthermore, DBMS is also responsible for managing the database users? requests [I-1]. Today, databases are considered essential for everyday organizations tasks. In healthcare units, due to hospital databases store very important information about the patients? clinical status, administrative information and other relevant information for the healthcare services, they are essential. Therefore, it is crucial to ensure the confidentiality, integrity and availability of data

[I-1][I-2][I-3].

However, database availability is a complex feature. There are several problems that influence database availability, such as: the database can be inaccessible due to network problems or a virus; the database can be too slow and therefore not satisfy the user?s requests [I-4].

The unavailability of the healtcare units databases is often related resource limitation faults. These faults occur when the physical resources are not enough to keep the database running. It is essential that the database must continue to operate despite the faults. For this reason, fault tolerant systems are already used. They are responsible to ensure the availability of data even a fault occurs. However, these systems do not allow the early detection of faults [I-5][I-6][I-7].

Therefore, the development of a fault forecasting system which allows taking early actions to solve problems is crucial. Forecasting models have been used in critical areas such as medicine. The objectives of this research is characterized the workload of the one database of Centro Hospitalar do Porto and a development of a model for forecasting and prevention of database faults by adapting an existing forecasting model in medicine (Modified Early Warning Score) [I-8].

The remaining paper is organized in the following sections: in the second section, we will address the issue of hospitalar information systems, showing the platform where our study was made ; in the third, we will present the forecast model (MEWS); At the fourth, will be described the methodology. In the fifth part, the results will be presented and discussed. At last, the seventh section are presented conclusions.

3.1.3 Hospital Information Systems

HIS can be defined as a subsystem hospital with a socio-technological development, which covers all hospital information processing[I-9]. Its main purpose is to contribute to the quality and efficiency of healthcare. This objective is primarily oriented to the patient after being directed to health professionals as well as the functions of management and administration [I-9][I-10]. The HIS also assumes much importance in relation to costs since the sector of communication technologies in healthcare is increasingly important [I-10]. There are four basic functional processes. This process list begins with the patient?s admission and ends in discharge or transfer to another institution. The other categories will serve to support the healthcare with the primary objective of improving quality. The four functional categories are characterized as follows: care; clinical process management; work organization and resource planning; and hospital management [I-11][I-12]. EHR can be assumed as a HIS for excellence and has replaced the traditional manual recording in Paper Clinical Process (PCP). EHR may include all hospital areas with a need for registration information. This information can be clinical, administrative and financial [I-13][I-14].

A. AIDA

AIDA means Agency for Integration, Diffusion and Archive of Medical Information. It is a platform that consists of a Multi-Agent System (MAS) and it can be considered like the main HIS where it has been working. The AIDA main goal is to overcome difficulties in achieving uniformity of clinical systems, as well as medical and administrative complexity of different hospital information sources [I-14]. AIDA was created by a group of researchers from the University of Minho and is currently installed in many Portuguese hospitals. It is an electronic platform that holds intelligence electronic employees (agents). This platform promotes a pro-active behaviour in its main functions: communication between hospital heterogeneous systems; storage and management all hospital information; response to requests in time; sending and receiving information from hospital sources like laboratories (medical reports, images, prescriptions). Thus, AIDA enables interoperability between hospital subsystems, assuming a main role where it is installed [I-15][I-16]. AIDA has an easy access for your users, allowing the management of clinical information anywhere in the hospital. In addition, the platform enables the sending of messages via phone or e-mail. The same way, AIDA establishes connection with all others systems of patients information: EHR; Administrative Information System (AIS) used by administrative people; Medical Information System (MIS) used in medical record; and Nursing Information System (NIS) used by nurses [I-17].

B. AIDA-PCE

The AIDA-PCE is an EHR and was implemented in the Centro Hospitalar do Porto. It is working as a subsystem of the main HIS. The AIDA-PCE follows a problem-oriented organization suggested by Lawrence Weed in the 60s. This information organization is known as the Problem Oriented Medical Record (POMR) and it assumes that registration is a production of clinical scientific document. In this kind of organization, the clinical information (annotations, therapeutics and treatments, diagnostics, diaries) should be recorded for specific problem, creating a list of issues organized in a tree structure, where each new problem derives from the main branch [I-13][I-18][I-19]. These problems must be classified as active or inactive, in which active problems are those where the disease is still active or even when medical intervention is required immediately. On the other hand, inactive problems require no urgent action. In this EHR problems assets are monitored and recorded daily using a SOAP (Subjective, Objective, Assessment and Planning) framework. Thus, each record contains the patient?s Symptoms, a doctor?s Observation, an Analysis of diagnosis and a treatment Plan that the patient is subject to [I-13][I-19].

The AIDA-PCE has many common features with PCP but it has a response, which is fast, reliable and safe. The structure of this EHR allows seamless integration with existing HIS by promoting the ubiquity of records between different specialties and services. The ubiquity of the AIDA-PCE allows access to mechanisms for monitoring alarm systems and decision support. The electronic record allows generation of documents and customized reports for specific purposes. It becomes easier to configure interfaces for registration and more. The information contained herein is standardized and uniform [I-13][I-15][I-20]. In the hospitals where AIDA and AIDA-PCE was installed, it was made a substantial investment to ensure the availability, reliability and scalability of the system.

C. AIDA and AIDA faults

As it was mentioned above, AIDA platform and AIDA-PCE are running at real behavior with real cases, because of this reason some problems have emerged. The problems are not responsibility of the AIDA and AIDA-PCE but these ones interfere directly with AIDA and AIDA-PCE operation and, consequently, the quality of the medical record and patients treatment. The main problems are due to communication faults in the network and are related to peak hours using it. Other situations are related to database faults, such as: bad management of datacenter; many users accessed at the same time; breaks on energy. Some of these crashes must to be avoided, but actually errors will always happen [I-1].

D. AIDA and AIDA-PCE fault tolerance mechanism

Fault tolerance mechanisms are used to ensure avaliability of data but also to load balancing and recovery [I-1][I-2]. AIDA and AIDA-PCE architecture, present in Figure 3.1, is composed by an Oracle Real Application Clusters System (RAC) and a dataguard solution. RAC is used for improving the availability and scalability.

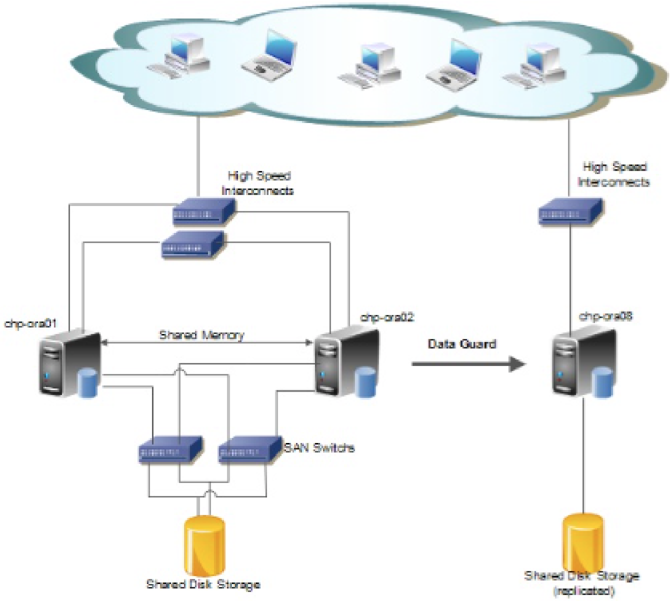

Figura 3.1: AIDA and AIDA-PCE Architecture

This architecture is composed by a shared repository of data which can be accessed through the server/computer that contains a database instance. If a server is down, it is possible access to the database easily by another server [I-1][I-6].

A data guard solution consists in one or more standby databases (replicas of the main database). When the main database is unavailable the replica can be used without the need to interrupt operation of the system. It is crucial that the main and the standby databases are synchronized and the access is read-only during recovering [I-2].

3.1.4 MEWS

The concept of Early Warning Score was been introduced in 1997 and the vital signs measured were: Temperature, Respiratory rate, Systolic BP, Heart rate and Neurological (**A**lert, Reacting to **V**oice, Reacting to **P**ain, **U**nresponsive). However, in 1999 a new model appeared with the use of two additional variants: urine output, saturation of hemoglobin with oxygen SPO2 [I-21][I-22]. This model is the Modified Early Warning Score (MEWS). MEWS assumes that a serious problem of health is often preceded by physiological deterioration.

This model consist in a strict and continuous monitoring of the patient's vital signs. Then, using the decision table, see Table 3.1, the scores are calculated to determine the level of risk of each patient, trying to understand when a serious problem will be occur [I-8][I-21] [I-22].

Normally, if any of the parameters have a score equal to two, the patient must be in observation. In the case of the sum of scores being equal to four, or there being an increase of two values the patient requires urgent medical attention. In a more extreme situation, if a patient has a score higher than four, his life is at risk. The use of this model allows identify the patients at risk and give them priority, improves the monitoring physiological parameters of the patients and thus support the medical decision [I-8][I-23].

Tabela 3.1: MEWS Score

Score	**3**	**2**	**1**	**0**	**1**	**2**	**3**
Temp (°C)		<35	35.1-36.0	36.1-38.0	38.1-38.5	>38.6	
HR (min -1)		<40	41-50	51-100	101-110	111-130	>130
Systolic (mmHg)	<70	71-80	81-100	101-199		> 200	
RR (min -1)		<8		8-14	15-20	21-29	>30
SPO2	<85	85-89	90-93	>94			
Urine (ml/Kg/h)	Nil	< 0.5					
Neurological				A	V	P	U

3.1.5 Methodology

A. Monitoring database performance

Monitoring is the first step for process of identify the source and symptoms of the faults. This is a complex process, however, Oracle provide several tools to help in this process. An example of these tools are the performance views which contain useful information for monitoring [I-24][I-25].

The most important views related to this paper are the ones which contain information about database component statistics. For global monitoring of the database must be used system-level views (v$sysstat, v$sys time model, v$system event). For more detailed information about the sessions should be used to session level views (v$sesstat,v$session event, v$sess time model [I-4][I-25]. It was also used an operating system command (*sar*) to gather relating to the memory and processor utilization [I-26].

There are others tools for database monitoring. However, the performance views were chosen for two reasons: they make the integration with the business intelligence tool used to do the analysis and presentation of data (Pentaho) easier; and it is easier to define the zone that is to be monitored, thereby reducing the cost associated with monitoring. A Java application was developed to monitoring AIDA-PCE database. According to the objective of preventing faults related to the resource limitation have been selected the following statistics [I-2][I-24][I-27]:

DB time - This statistic gives information related to database response time. The response time is the period between an initial user request and the return of the results. In Oracle systems, this time is a sum of total time (including CPU time, IO time, Wait time). Therefore it is a good indicator of the workload of the system. Typically, this time increases with the number of simultaneous users or applications, but it also may increase due to large transactions [I-2][I-28].

Number of transactions - Database had more or less work according to the transactions which are performed. In Oracle databases, the number of transactions can be obtained by adding up the values of statistics“user commits” and “user rollbacks” due each transaction always ends with a “commit” command and any undo operation as a “rollback” command [I-2][I-4][I-26].

Number of Operations - One transaction may trigger a large set of operations (sum of user calls and recursive calls) depending the query. This means that there may be a large number of transactions and a low number of operations or otherwise. In Oracle databases this information can be obtained by collecting, the "execute count" statistic [I-26].

Number of sessions - The collection of this statistic is important because each session is associated with a piece of memory, so many simultaneous sessions can cause problems. In Oracle systems, the number of sessions can be obtained by statistic "logons current" [I-24].

Processor utilization - The processor is one of the most important components, so it is necessary to constantly monitor its utilization by the user processes. Low values of processor utilization may indicate problems at the level of I/O. If the values are too high, it can compromise the functioning of the database. The percentage of utilization can be obtained thought a command of operating system such as sar -u[I-24].

Memory utilization - The memory is a key component to the speed of the database systems. The speed of access to data depends on the place where they are: memory or disk. If the data are in memory then access to them is faster. This statistic is also accessible through the operating system commands such as sar -r [I-4].

Size of redo file - The redo files are used to store information about changes made to the database. These are very important for the recovery of faults. An increase in the size of these files, it indicates a higher number of operations and therefore a higher database load. In Oracle systems, the size (kb) of redo file can be obtained by statistic "redo size" [I-25].

Amount of I/O requests - The I/O operations are very time consuming often are associated with writing or reading data from memory to disk. An excess of this kind of operations can indicate problems in memory [I-24]. Amount of redo space requests - Indicates the lack of space to write in the redo buffer, this can lead to delays because it is necessary to write some data to disk to release memory. This can happen due to a poorly sized buffer, or excess entries generated simultaneously [I-29].

Volume of network traffic - Network is very important for database performance because several database components are connected through network, and all the user requests come from the network. If a volume there is a problem in network, the database can be slow and compromise users' requests [I-24].

Recursive Calls ratio - Calls to database can be of two types: user calls or recursive calls. When a user request can be resolved through a single SQL query, this is a call. A recursive call occurs when a user request need one query SQL that needs another SQL query. Ideally this ratio should be as low as possible. The recursive calls ratio is the fraction between the recursive calls and total calls (user calls + recursive calls). The high value can indicate problems with the design of tables or an excessive amount of triggers running at the same time. This ratio can be calculated by the equation [I-25][I-29].

Buffer cache ratio (BC) - This ratio shows the percentage of data that is in memory cache, rather than in the disk. Normally, the BC is very high so it is necessary to pay attention if BC decreases, this may indicate lack of memory problems. BC can be calculate by the the fraction between the number of disk accesses (physical reads) and the number of memory accesses (consistente gets + block gets).

The monitoring program collected values during a month. Using a Business Intelligence tool it is possible analyze the data collected and create graphs that demonstrate the normal behavior of the database. For each statistic, were calculated the upper (using the 75th percentile) and lower (using the 25th percentile) limits.

B. Business Intelligence (BI) tool

BI tools are useful because database monitoring provides a large quantity of data which is difficult to analyze. Thus, BI tools allow the management of the data and present the conclusions of a more clearly way.

The BI tool chosen for this research was Pentaho Community Version because it has all the desired features. The Pentaho Community version provides tools that allow the creation of reports, dashboards/charts, application of data mining, integration techniques and data modeling [I-30][I-31]. There is a main application, the bi-server, where it is possible to perform reporting and analysis. In this application, a plugin which allows the development of dashboards can be added. However, other applications that can be easily integrated with the main server in order to add new features. The applications used in this paper are [I-31]:

- **CDE** - Allows the user to develop the updateable dashboards, in order to show more clearly the data collected/treated. It was developed and maintained by Web-details. The front-end is based on HTML, and dashboards can be populated with data coming from a variety of sources, such: SQL querys, xml files, mondrian cubes, spoon transformations [I-31].
- **Pentaho Data Integration (Spoon)** - Delivers powerful Extraction, Transformation and Loading (ETL) capabilities. This tool is very useful to integrate information from different sources and also to perform some mathematical operations on data. The division of a multi-step transformation reduces the complexity of the SQL query used to obtain the desired information [I-31].

After represented the normal behavior of the database, a new program was developed for detecting the abnormal situations. Depending on the value of the deviation, for each statistic, is assigned scores to abnormal situations such as is done in MEWS. Two situations can happen (see Table 3.2): if the score is greater than 0 and less than 3, a visual warning will be issued on the dashboard; if the score is greater than 3, warnings will be sent via email to the database administrator, allowing he to take speedy action to prevent the occurrence of a fault in the database. This program is upgradeable, as new limits calculated at the end of the day based on new measurements, because it is possible that exists an increase of load and the database remains operational.

Tabela 3.2: Severity Score

Score	**0**	**1**	**2**	**3**
Value	$< p75$	$< p75 < p80$	$< p80 < p90$	$> p90$
Severity	Normal	Low severity	Severe	Critical

3.1.6 Results

As a first step will be examined workload of the database. In Figure 3.2, is represented a table that contain the mean values of each statistical measure for each node. It is possible observe that node 1 (chp-ora01) has more workload than node2 (chp-ora02). Only in the ?redo log space requests per second?, the mean of node1 is not higher to the node 2.

Overall, the AIDA database has an average 674 sessions and about 290 transactions per second, which shows that is a database with a high workload. The high value of use of the network, shown that it is also a very important characteristic for evaluating the performance of the database. Concerning the use of two main components, memory and processor, through the Figure 3.3 is possible to observe that the average is practically the same in both nodes. Memory presents high values unlike the processor. This may be due to take into account only the processor utilization by user processes. To identify the most critical points of the day, graphs and tables of Figures 3.4, 3.5 were constructed. In Figure 3.4, it is possible identify that the period between 11:00 and 12:00 is the most critical period in node1. In this

Show 10 entries Search:

Name	Average (node1)	Average (node2)	Total (node1+node2)	Comparison (nodo1-nodo2)
DB time per second	3.015	2.987	6.002	
Network Traffic Volume (bytes/sec)	380657.63	373348.54	754006.17	
Number of operations per second	414.634	404.997	819.631	
Number of I/O requests per second	376.897	365.629	742.526	
Number of sessions	340.554	333.892	674.446	
redo log space requests per second	0.005	0.005	0.01	
Number of redo size (kb/s)	106909.058	49782.597	156691.655	
Transactions per second	146.136	144.424	290.56	

Showing 1 to 8 of 8 entries

Figura 3.2: AIDA Workload

period, there is three important peaks: DB time, percentage of processor usage and number of sessions.

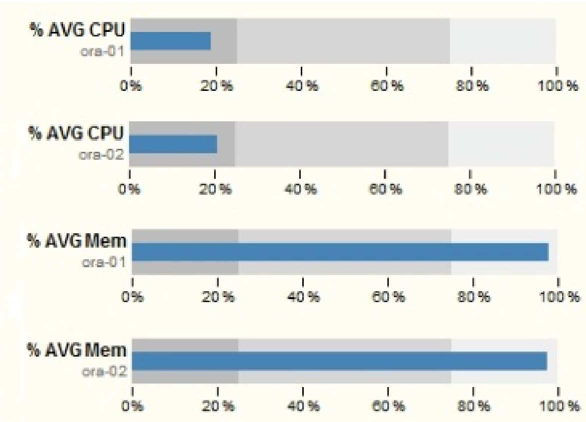

Figura 3.3: AIDA Workload

Figure 5, presents the workload peaks for node two (chp-ora02). In this node, the distribution of peaks is more diverse. However, in the period between 10:00 and 11:00, there is three important peaks: Network traffic, percentage de processor usage and recursive call ratio. In general one can consider the period from 10:00 to 12:00, as the period most favorable for the occurrence of faults. For each statistic of each node limits were calculated by using percentiles. The Figure 3.6, shows the normal

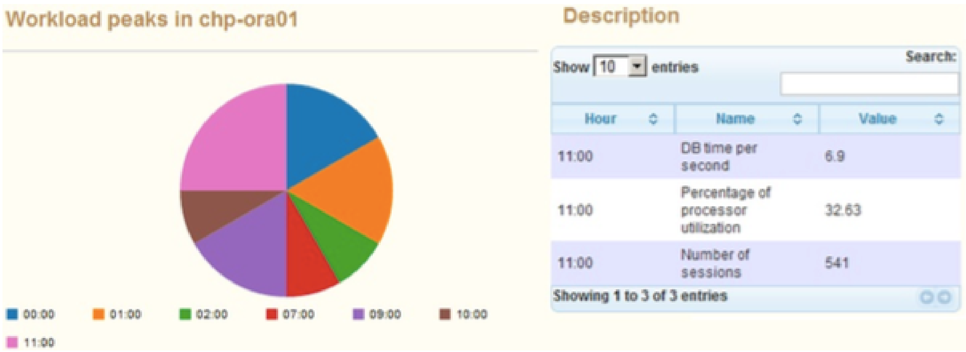

Figura 3.4: Workload peaks chp-ora01 node

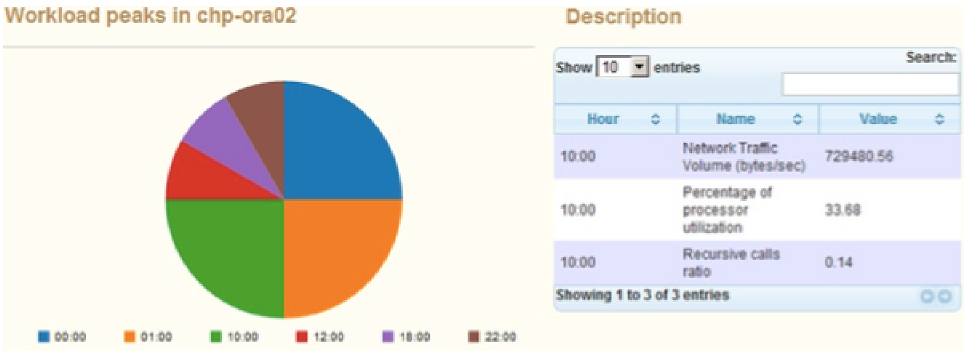

Figura 3.5: Workload peaks chp-ora02 node

behavior of the variable “number operations per second” throughout the day. For this example only the period between 08:00 and 19:00 is considered because in this range there is the highest number operations per second. It is possible to see that the database has more activity in the morning, due to the higher number of users connected. It is expected that 50% of the collected data are present between the two limits defined by the 25th and 75th percentiles. In order to identify the abnormal situations scores were assigned to percentiles. The data above the percentile 75 is considered a low severity situation, above 80 a grave situation and above 90 a critical situation (see Table 3.2). Figure 3.7 presents abnormal situations that occur along one day. In the period between 12:00 and 14:00 there is two abnormal situations. The first is a severe situation and the other is a critical situation, according to Table 3.2. However, these situations do not cause database fault. For this reason it is necessary to update the limits. Limits are update in the end of day taking account of all measured values which do not cause fault.

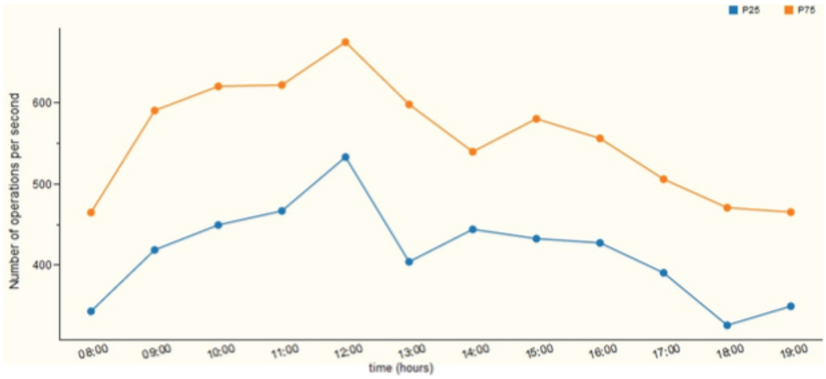

Figura 3.6: Limits of number of operations per second in chp-ora01

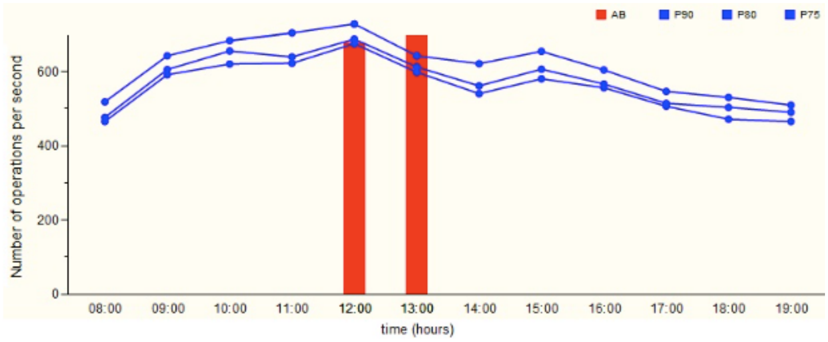

Figura 3.7: Limits and abnormal situations in chp-ora01

3.1.7 Conclusions

A model was proposed to characterize and represent the normal workload for AIDA database. It was observed that the node1 of the database is more used than node2. Moreover, workload peaks were identified and it was possible to observe that the critical period is the period between 10:00 am and 12:00 am.

With percentiles, it was possible to determine the limits that characterize the abnormal situations. A decision table, with percentiles and scores was created. Therefore, the methodology of MEWS were adapted to database reality through an upgradeable and learning forecasting fault system. The Pentaho Community allowed to perform the data analysis and development of dashboards which makes it easy the interpretation of results these results.

Due to the heterogeneity of the database workload it was found that the limits will be updated every day.

3.1.8 References

[I-1] S. Drake, W. Hu, D. McInnis, M. Skold, A. Srivastava, L. Thalmann, M. Tikkanen, O. y. Torbjornsen, and A. Wolski, “Architecture of Highly Available Databases,” in Service Availability, ser. Lecture Notes in Computer Science, M. Malek, M. Reitenspie, and J. Kaiser, Eds. Springer Berlin / Heidelberg, 2005, vol. 3335, pp. 1-16.

[I-2] R. Godinho, “Availability, Reliability and Scalability in Database Architecture,” MSc Thesis, Universidade do Minho, 2011.

[I-3] A. Rodrigues, Oracle 10g e 9i:Essentials For Professionals (in portuguese). Lisboa: FCA, 2005.

[I-4] R. Schumacher, Oracle Performance Troubleshooting With Dictionary Internals SQL & Tuning Scripts. Kittrell: Rampant TechPress, 2003.

[I-5] O. Yi, “Highly Available Database Systems,” in Integrated information systems and databases seminar, 2006.

[I-6] M. Cyran, P. Lane, and J. Polk, Oracle Database Concepts, 10g Release 2 (10.2). Oracle, 2005. [Online]. Available: http://docs.oracle.com/cd/B19306 01/server.102/b14220.pdf

[I-7] W. Hodak, S. Kumar, and A. Ray, “Oracle Database 11g High Availability,” Oracle, Tech. Rep., 2007.

[I-8] C. P. Subbe, M. Kruger, P. Rutherford, and L. Gemmel, “Validation of a modified Early Warning Score in medical admissions,” QJM, vol. 94, no. 10, pp. 521-526, 2001.

[I-9] R.Haux,A.Winter,E.Ammenwerth,andB.Brigl,StrategicInformation Management in Hospitals: An Introduction to Hospital Information Systems. Springer-Verlag, 2004.

[I-10] J. Machado, A. Abelha, P. Novais, J. Neves, and J. a. Neves, "Quality of service in healthcare units," Int. J. Computer Aided Engineering and Technology, vol. 2, no. 4, pp. 436-439, 2010.

[I-11] J. Duarte, C. F. Portela, A. Abelha, J. Machado, and M. F. Santos, "Electronic health record in dermatology service," in Communications in Computer and Information Science, 221 CCIS (PART 3), 2011.

[I-12] H. AmmenwerthE., BuchauerA., "Arequirementsindexfor information processing in hospitals," Methods of Information in Medicine, vol. 41, pp. 282?288, 2002.

[I-13] M.Miranda,G.Pontes,P.Gonçalves,H.Peixoto,M.Santos,A.Abelha, and J. Machado, "Modelling intelligent behaviours in multi-agent based hl7 services," in Studies in Computational Intelligence, Springer, Ed., vol. 317, 2010.

[I-14] E. Coiera, Guide to Health Informatics (2nd ed.), 2nd ed. London: Hodder Arnold, 2003.

[I-15] J. Duarte, J. Neves, A. Cabral, M. Gomes, V. Marques, M. F. Santos, A. Abelha, and J. Machado, "Towards intelligent drug electronic prescription," in European Simulation and Modelling Conference, Guimarães, Portugal, 2011.

[I-16] F. Portela, M. Vilas-Boas, M. F. Santos, A. Abelha, J. Machado, A. Cabral, and I. Aragao, "Electronic health records in the emergency room," in ACIS-ICIS 2010, pp. 195-200.

[I-17] V. Slee, D. Slee, and J. Schmidt, The Endangered Medical Record: Ensuring Its Integrity in the Age of Informatics. Saint Paul, Minnesota: Tringa Press, 2000.

[I-18] M. Miranda, J. Duarte, A. Abelha, J. Machado, and J. Neves, "Interoperability and healthcare," in European Simulation and Modelling Conference, Leicester, UK, 2009.

[I-19] K. Hyrinen, K. Saranto, and P. Nyknen, “Definition, structure, content, use and impacts of electronic health records: A review of the research literature,” International Journal of Medical Informatics, vol. 77, pp. 291-304, 2008.

[I-20] C. Bossen, “Evaluation of a computerized problem-oriented medical record in a hospital department: Does it support daily clinical practice?” International Journal of Medical Informatics, vol. 77, pp. 592-600, 2007.

[I-21] A. Albino and V. Jacinto, “Implementation of early warning score - MEWS,” Centro Hospitalar do Barlavento Algarvio, EPE, Portimão (in po?rtuguese), Tech. Rep., 2009.

[I-22] J. Gardner-Thorpe, N. Love, J. Wrightson, S. Walsh, and N. Keeling, “The value of Modified Early Warning Score (MEWS) in surgical inpatients: a prospective observational study.” Annals of the Royal College of Surgeons of England, vol. 88, no. 6, pp. 571-5, Oct. 2006.

[I-23] G. Devaney and W. Lead, “Guideline for the use of the modified early warning score (MEWS),” Outer North East London Community Services, Tech. Rep., 2011.

[I-24] I. Chan, Oracle Database Performance Tuning Guide, 10g Release 2 (10.2). Oracle, 2008.

[I-25] K. Rich, Oracle Database Reference, 10g Release 2 (10.2). Oracle, 2009.

[I-26] C. Shallahamer, Forecasting Oracle Performance. Berkeley: Apress, 2007.

[I-27] L. Ramos, “Performance Analysis of a Database Caching System In a Grid Environment,” MSc Thesis (in portuguese), FEUP, 2007.

[I-28] K. Dias, M. Ramacher, U. Shaft, V. Venkataramani, and G. Wood, “Automatic performance diagnosis and tuning in Oracle,” in Proceedings of the

2005 CIDR Conf, 2005.

[I-29] Hoopoes, "Oracle Database Tuning Statistics," 2007. [Online]. Available: http://www.hoopoes.com/cs/oracle tune.shtml

[I-30] M. Tereso and J. Bernardino, "Open source business intelligence tools for SMEs," in Information Systems and Technologies (CISTI), 2011 6th Iberian Conference on, 2011, pp. 1-4.

[I-31] Pentaho, "Welcome to the Pentaho Community." [Online]. Available: http://community.pentaho.com/

3.2 II - Step Towards Fault Forecasting in Hospital Information Systems

Paulo Silva, César Quintas, Manuel Santos, António Abelha and José Machado

5th IEEE International Conference on Biomedical Engineering and Informatics, Chongqing, China, 2012.

3.2.1 Abstract

Nowadays, many organizations consider databases indispensable tools for their daily tasks. Particularly in healthcare units, databases have a vital role, since they archive very important information about patients? clinical status, therefore, it is crucial that databases are available twenty-four hours a day, seven days per week. Healthcare units have already implemented fault tolerant systems, which intended to ensure the availability, reliability and disaster recovery of data. However, these mechanisms do not allow to take preventive actions in order to avoid the occurrence of faults. In this context, the necessity of the development of faults prevention and prediction systems is emerging. These systems can predict faults with some time in advance and provide taking early action to solve problems. The objectives of this paper are: monitor database performance and adapt a forecasting model used in medicine (Modified Early Warning Score - MEWS) to database reality.

3.2.2 Introduction

Databases are powerful tools to record and manage of the large amount of organizations' data. They have several components since software components (application, management) until hardware components (disks, memory, processor, connectors) that are coordinated by the Database Management System (DBMS). Besides these components the DMBS has to take into account the users and their satisfaction, to perform their requests in time [II-1][II-2][II-3].

Actually, DBMS solutions provide by Oracle are one of the most well know and most used solutions for database management. Over the years, organizations have increased the use of databases and DBMS, which, for simplicity, sometimes are all denominated only by database [II-3]. Today, many organizations consider databases indispensable tools for daily tasks [II-4]. Particularly in healthcare units, databases have a vital role, since they archive very important information about patients?

clinical status and other information relevant to proper functioning of the healthcare unit. So it is crucial that these databases provide a high level of security to ensure permanent data integrity and availability. One of the fundamental characteristics of database security is the availability. In these units, it is very important that databases are available twenty-four hours a day, seven days per week. Therefore, database availability is a complex and crucial feature [II-3].

Database availability is a complex question, because the fact of the database is "up" is not a synonym that it is available. There are several problems that influence database availability, such as: database can be inaccessible due to network problems or a virus that bar user's connections; the database can be too slowly and therefore the users' requests are not satisfied. The slowness of database can be related to resource limitation or an elevate number of operations. Database has intermittent faults and loose the user's confidence, because of its unavailability [II-5][II-6]. Database faults can happen frequently, and can be of several types depending of the vendor and environment where database is installed. Nevertheless, it is very important that these faults are transparent for users, so that database continues available despite faults [II-7]. Healthcare units have already implemented some fault tolerant systems, which intend to ensure the availability of data when a fault occurs. However, it is very important the development of fault's forecasting and prevention systems, which allow to taking early actions to solve future problems by identified the abnormal situations. An alternative to these mechanisms can be the forecasting models that have been used in critical areas such as medicine [II-8].

The objective of this paper is the development of a model for forecasting and prevention of database faults by adapting an existent forecasting model in medicine (Modified Early Warning Score - MEWS) [II-8]. A monitoring program was developed and with business intelligence tools knowledge was extracted in order to the study of the data. This is a base for the development of a forecasting fault application.

3.2.3 Health Information Systems

Over the years, healthcare units have increased the utilization of computer systems in many of their services in order to introduce new methodologies for problem solving [II-9][II-10]. In this sense, healthcare information systems have emerged. These systems are responsible for optimizing the set of information which exists in a healthcare unit. They proceed to the collection, processing, and data management of all stakeholders (patients, nurses, doctors) and of all services of healthcare unit. Thus, in a particular situation, the authorized people have a fast and efficient access to relevant information [II-9]. Rapidly the information systems have taken a key role in healthcare units. Since 1990s, hospital information systems, in Portugal was implemented the Hospital Information Integrated System (SONHO). This system had as objectives merge clinical and administrative information. The evolution of these systems is the genesis of the creation of the electronic medical record, and of the addiction of two modules: the medical support system (SAM) related to medicine and the nurse support system (SAPE), related with nursing care [II-11].

However, in healthcare units there is a lot of other large information systems, they are distributed and heterogeneous systems and in order to communicate an effort was made to develop interoperability [II-12]. To solve this problem a dynamic framework (AIDA- Agency for the Integration, Diffusion and Archive) was developed. This framework is composed of pro-active agents that are responsible for promoting communication by sending/receiving information, and managing and saving the information. Thus, AIDA shares information and knowledge among every information systems, namely: the administrative information system (AIS); the medical support information system (MIS); the nursing support information system (NIS); the EMR information system; the department information systems (DIS) of all the departments or services [II-12][II-13]. The information is archived in very large databases which must be available every day of the year because their information is vital for solving the patients' problems and for hospital management.

Therefore, it is crucial to ensure the integrity and permanent availability of data in case of faults [II-4]. To achieve this goal it is necessary to use tolerance fault mechanisms, which determines data redundancy, component redundancy or both [II-3]. Through this process of redundancy, it is possible for a database to still be available in spite of failures in the hardware or software components of the system. Some mechanisms also allow for load balancing and recovery in extreme situations, such as fire [II-4][II-14].

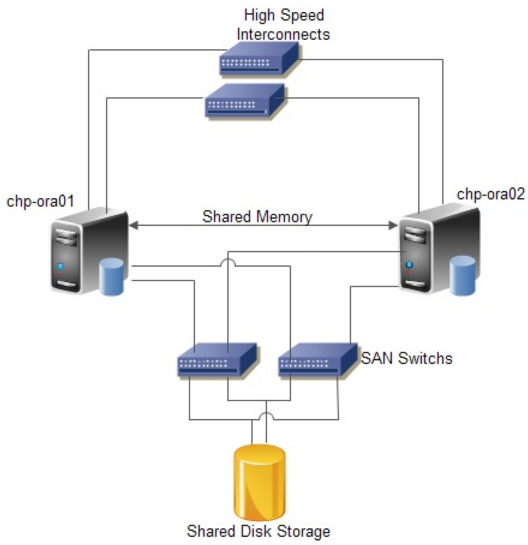

Figura 3.8: AIDA RAC Architecture

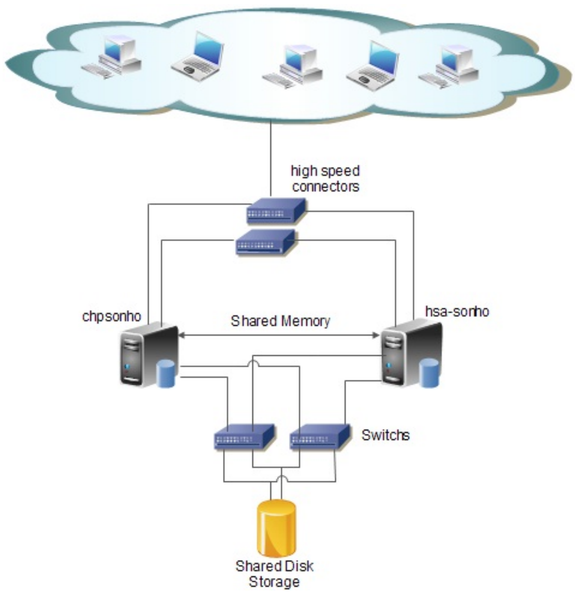

Figura 3.9: SONHO RAC Architecture

In the particular case of the AIDA and SONHO databases of Centro Hospitalar do Porto are based in an Oracle Real Application Cluster system (RAC). This mechanism is provided by Oracle for improving the availability and scalability of databases. This goal is achieved through architectures presented in Figures 3.8 and 3.9. This architecture is composed by a shared database which can be accessed through the server/computer that contains a database instance and an Automatic Storage Management (ASM) instance [II-3][II-15][II-16]. In addition to the RAC system there is one data guard solution (Figure 3.10). A data guard solution consists in one or more standby databases (replicas of the original database) which should be in different places. Thus, if for some reason the main database is unavailable, the system will still work by using one of the standby databases. It is important that the main and the standby databases are synchronized and the access is read-only during recovering [II-4]. In spite of all advantages, the fault tolerance tools do not allow the focus on the faults themselves, but only on the faults' effects, trying to minimize them. Thus, it is necessary to monitor the components of the database, trying to understand their faults and symptoms.

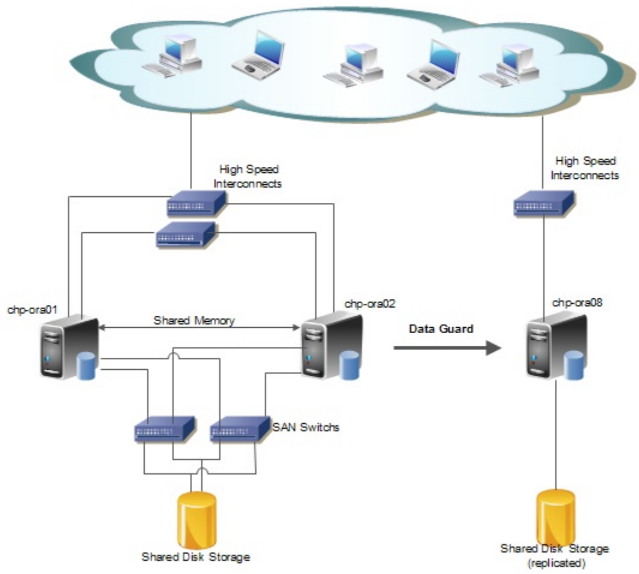

Figura 3.10: AIDA Complete Architecture

The AIDA and SONHO databases are large databases which a large quantity of information: AIDA includes around 5451 tables, SONHO has around 2255 tables and with a large number of sessions and processes that run permanently. Therefore, most often database faults are related to resource database limitation. So, the focus of this paper is in forecasting fault database related with database resources, adapting the forecasting model used in Medicine (MEWS) of the world of databases.

3.2.4 MEWS

In medicine a model for the prediction, in advance, of serious problems of health is already used. This model is Modified Early Warning Score (MEWS) and it assumes that a serious problem of health is, sometimes, preceded by physiological deterioration. So it is an important procedure to monitoring patient?s vital signs that depends on the deviation from normal assigned scores. In Table I, it is possible to find the vital signals and the score associated to each set of values.

These scores are added and it is determining the level of risk of each patient, trying understanding when a serious problem occurs [II-8][II-17]. This model is very important because the resources of intensive care units are limited, and then it is possible to identify patients at risk [II-17]. The concept of Early Warning Score was been introduced in 1997 and the vital measured signs were: temperature, systolic blood pressure, respiratory rate, heart rate and neurological activity [II-17].

However, some years later several studies have concluded that more parameters for the success of this model were needed. So, in 1999 a new model appeared with the use of two additional variants: urine output and saturation of hemoglobin with Oxygen (SPO2) [II-17][II-18][II-19]. The monitoring of patient?s vital signs should be continuous and all values must be archived to understand the behavior of the vital signs over time [II-18][II-20].

Normally, if any of the parameters have the score equal to two, the patient must be in observation. In the case of the sum of scores is equal to four, or there is an increase of two values the patient requires urgent medical attention. In a more extreme situation, if a patient has a score higger than 4, is life is in risk [II-8][II-20].

The use of this model has a lot of advantages in the process of preventing patient?s medical complications, such as:

- Allow defining priority for the medical interventions [II-17].
- Improve the monitoring and observation of the patients since it gives an indication of physiological parameters [II-20].
- Support the medical decision, it is more objective because it is based on quantitative criteria [II-18].

3.2.5 Methodology

A. Monitoring database performance

The first step to construct a forecasting database's fault system is monitoring database components that can cause faults. The monitoring database components is not only for fault diagnostic but also for improvement of security because with database monitoring it is possible, for example, to identify an abnormal resource utilization for an unknown user. Therefore, the use of monitoring tools has increase in organizations [II-21]. The monitoring difficulty increases with the dimension and complexity of databases. Normally, hospital's database had several components and complex architectures. Therefore, hospital database monitoring is not a simple task; however, the Oracle systems provide several tools to help in this process. A sample of these tools is the performance views (v$). These are a set of views where useful information for monitoring process is available. A view looks like a simple database table, but since views show the actual system state, it is impossible do editing operations, is only possible to consult information. These views are sometimes called dynamic performance views because their content is refreshed several times during a database instances' life [II-22] [II-23].

The most important views related to this paper are the ones which contain information about database component statistics, such as processor, memory, disk, sessions and users. These statistics can be divided into three main groups [II-22][II-24]:

- **Time Model Statistics** - Provides information about the time spent in the database calls.

- **Wait Event Statistics** - Provides information about the time the server process/thread had to wait for an event to complete before being able to continue processing.

- **Session and System Statistics** - Provides information about the use of system resources, about users and their sessions.

However, this tool is not enough. It is necessary to use others tools more related with physical resources. Usually commands of operating system, which in this case are UNIX command, are used. There are several commands that can help in the monitoring process. For example, we can use the command **top** or the command **sar** to monitor processor (CPU) and memory usage, and we can use the command **ps** to see the active process in the moment. Using these tools together we can do a more efficient monitoring [II-25].

There are others tools for database monitoring such as Automatic Workload Repository (AWR), or Nagios [II-22]. However, it is chosen the performance views for two reasons: it provides a easy connection with the Business intelligence tool (Pentaho) for presentation and data analysis; and it is easier to define the zone of monitoring thus reducing the cost associated to monitoring.

B. Business Intelligence(BI) tool

In this paper, the use of BI tools is useful because database monitoring provides a large quantity of data which are difficult to interpret and manipulate.

The term Business Intelligence (BI) was introduced by Howard Dresner of the Gartner Group to describe a set of concepts/methods which utilize new technologies to improve the business on decision making [II-26] [II-27]. Over the years, many different BI definitions have appeared depending on the approach (managerial, technical, system-enabler) followed by each author [II-26].

In general, BI is related to a set of concepts, methods, processes, and tools used to improve business, particularly, decision-making. To achieve this goal, several steps are necessary. First, it is necessary to acquire and integrate data from different sources and archives. This data are stored in a large database (data warehouse). Then it becomes possible to perform several kinds of analyses to obtain useful information for decision making and reasoning for past experiences which allows for a precise understanding of business dynamics [II-26][II-28][II-29]. Implementation of these techniques is made by several analysis tools, such as production reporting, end-user query, data mining, dashboard, On-Line Analysis Processing (OLAP) and planning tools [II-26]. The utilization of BI in organizations brings many benefits, such as cost-saving in data consolidation, time savings for user requests, more and better information, better decisions, and more support for the accomplishment of strategic business objectives [II-28].

After reviewing several studies about BI open-source tools, it was observed that Pentaho Community version is more appropriate to accomplish this work's objectives [II-30] [II-31]. The Pentaho tool was released in 2004 having two versions: a commercial version (Enterprise Edition) and a free version (Community Edition). The community version provides several features, such as reporting, dashboards, charts, data mining tools and data exportation. This tool consists of several java applications which means that it can be used on different computational platforms. There is a main application, the bi-platform where applications can be integrated in order to turn available additional functionalities [II-30] [II-32]. The mentioned applications are [II-32]:

- **CDF-CDE** - Allows the user to develop dashboards, in order to show more clearly the data collected/treated.
- **Pentaho Data Integration (Spoon)** - Delivers a powerful Extraction, Transformation and Loading (ETL) capabilities.
- **Mondrian** - Allows business users to analyze large quantities of data in real-time (OLAP).
- **Pentaho Report Designer** - Allows the user to create relational and analytical reports from a wide range of data-sources, in several output files.
- **Weka** - Used to apply machine learning and data mining techniques. Weka is an independent free program which can be integrated with Pentaho by Weka Scoring Plugin, in the Pentaho Data Integration application.

In this paper, it is only used Pentaho Data Integration for manipulation of data and Community Dashboard Editor for developing charts and dashboards.

C. Explain of the statistics

Several statistics can be used for determining normal database performance. We developed a forecasting fault model for the database workload. The chosen statistics are related with the load of databases. The chosen statistics are: DB time, number of transactions, number of executions, calls ratio, number of current logons, processor and memory utilization, size of redo, buffer cache ratio, amount of I/O requests, amount of redo space requests and network traffic [II-4] [II-24] [II-33].

- **DB time** - This statistic gives information about database response time. The response time is the period between an initial user request until the return results. This time should be minimized. In Oracle systems, this time is a sum of total time (include CPU time, IO time, Wait time) spent on all requests from users. Therefore it is a good indicator of the workload of the system as it usually increases with the number of users and number of applications, and sometimes can also increase due to the larger transaction, or when there is some problem in the system [II-4] [II-34] [II-35].

- **Number of transactions** - In databases, transactions are considered units of work. In Oracle databases, the number of transactions can be obtained by summing the values of statistics "user commits" and "user rollbacks" because each transaction always ended with a command "commit" and any undo operation as a command "rollback" [II-4] [II-5] [II-25].

- **Number of executions** - One transaction may trigger a set of operations in database depending of query and information needs. Therefore, it is important to collect information about the number of operations. In Oracle databases this information can be obtained collect, the "execute count" statistic [II-25].

- **Calls ratio** - One transaction results in several calls. These calls can be of two types: user calls or recursive calls. When, after a request from a user, only one SQL statement runs is considered an user call. If the statement SQL requested by the user, need another SQL statement to meet the request of the same, then there is a recursive call. This ratio can be calculated using statistics "user calls" and "recursive calls". Ideally, this ratio is as low as possible, because a large number of recursive calls may indicate: problems with the design of tables, an increase of sorts because of disk cache memory problems, or excessive running trigers [II-23] [II-36].

- **Number of current logons** - The peaks of users are not directly related to the users, but with the number of sessions that may be simultaneously using the database, since a user can have more than one session. This value can be obtained through the statistical "logins current". It is important to determine the time of day, that have a higher number of sessions since each session is assigned to a portion of memory (PGA), which can influence the database performance [II-24].

- **Processor utilization** - It is important to analyze the amount of processor (CPU) used, because if the processor is in little-used may indicate that there is a high level drive I/O, which may indicate a problem. Other behavior indicative of problems, is the case in where an CPU is widely used but only for a small number of programs, which means that some processes never never have access to the processor. The observation of the host CPU, in percentage, can be obtained by operating system commands [II-24].

- **Memory utilization** - The memory is a key component to the speed of the database systems, since, if the data stays in memory, it improves the response time. However, if memory is busy, response time tends to decrease. Therefore it is important to determine when there are peaks of memory usage since they can lead to a problem of performance of the system [II-5].

- **Size of redo file** - Represents the amount of redo entries (kb). It is very important due to the function of redo files (archive the modifications in database). An increase of this statistic value can indicate an increase of the number of operations, therefore, the database load also increases [II-23].

- **Buffer cache ratio** - This ratio show the percentage of data that is in memory cache, rather than disk. This ratio can be calculate using the equation:

$$BC = \frac{1 - (physicalreads)}{(consistentegets + blockgets)} \tag{3.1}$$

However, care must be taken because if is used cumulative values, the ratio does not show any changes and there may be some error, although it can not be seen because the variables' magnitude [II-5] [II-35].

- **Amount of I/O requests** - The I/O operations demand a lot of time, and an excess of this kind of operations can indicate problems in memory. If there is no free space in memory, it is necessary to put data on the disk, which can damage the performance of the database [II-24].

- **Amount of redo space requests** - When there is no space in Redo log buffer more space is requested to put redo entries. This can happen due to small size of redo log buffer or due to an abnormal load of database that increases the number of operations [II-36].

- **Volume of network traffic** - The several database components are often connected through network and the user requests come from network. Therefore, the network is also important for database performance. If a volume of network is increasing rapidly, the database can be slow and compromise users' requests, so it is a very important statistic [II-22].

D. Process

Initially, a java application was developed to collect data related with statistics described above from the two major databases (AIDA, SONHO) of Centro Hopitalar do Porto. The application collected information during a week. With this information it was possible to make a model that represents a normal behavior of workload database. A number of charts and dashboards were created, providing important conclusions about the performance, normal behavior and utilization peaks of the database system.Using percentiles, some limits were defined to represent intervals of normal behavior. Thus, it is possible to detect abnormal situations as those statistics on the measured value is not within limits.

A new program was developed for catching the abnormal situations. Depending on the value of the deviation and the frequency of times that happen abnormal situations are assigned granted scores like it is done in MEWS. According to these scores, alerts will be sent to database administrator. This program is upgradeable because it calculates new limits when a new collect has done, using values within the limits, so the program continuously learn.

3.2.6 Results

This program is based in learning rules. Dynamically it creates a new model from normal behavior. It is presented an example for one database. In Figure 3.11, it is represented a limit of a normal number of transactions per second in SONHO database along the day is represented.

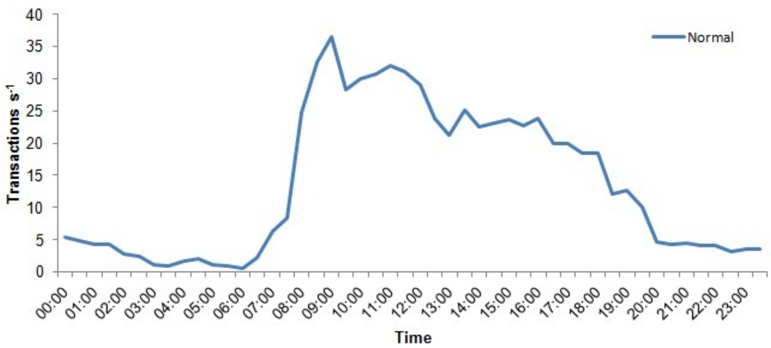

Figura 3.11: Transactions per second in SONHO

It is possible to see that the database has more transactions activity during morning, due to higher number of users connected. In Figure 3.12, abnormal situations that occur along one day are represented. The abnormal situations occur at the end of the morning. There is a repetition of abnormal values, this is a serious situation, so the score is 3. A warning of possible fault should be issued. However, database continues in operation despite the alerts.

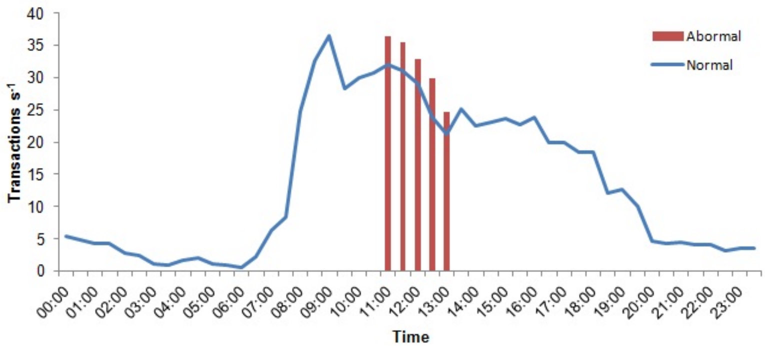

Figura 3.12: Abnormal situations: Transactions per second in SONHO

3.2.7 Conclusions

This paper shows how important is the theme of faults forecasting in healthcare databases. Despite being a complex theme, we propose a model to represent the normal workload for databases AIDA and SONHO. With this model, it is possible to do several analysis and with help of other mathematical tools it is possible to develop a program that detects the abnormal situations. According to the frequency and intensity of these abnormal situations it is possible to create a decision table with scores that represent the situation gravity. The scores represent the situation gravity. Thus, the MEWS were adapt to database reality through an upgradeable and learning forecasting fault system.

3.2.8 References

[II-1] T. Connolly and C. Begg, Database Systems A Pratical Approach to Design, Implementation, and Management, 4th ed. Harlow: Addison- Wesley, 2005.

[II-2] G. W. Hansen and J. V. Hansen, Database Management and Design, 2nd ed. Upper Saddle River, N.J.: Prentice Hall, 1996.

[II-3] S. Drake, W. Hu, D. McInnis, M. Skold, A. Srivastava, L. Thalmann, M. Tikkanen, O. y. Torbjornsen, and A. Wolski, "Architecture of Highly Available Databases," in Service Availability, ser. Lecture Notes in Computer Science, M.

Malek, M. Reitenspie, and J. Kaiser, Eds. Springer Berlin / Heidelberg, 2005, vol. 3335, pp. 1-16. [[II-Online]. Available: http://dx.doi.org/10.1007/978-3-540-30225-4 1

[II-II-4] R. Godinho, "Availability, Reliability and Scalability in Database Architecture," MSc Thesis, Universidade do Minho, 2011.

[II-II-5] R. Schumacher, Oracle Performance Troubleshooting With Dictionary Internals SQL & Tuning Scripts. Kittrell: Rampant TechPress, 2003.

[II-II-6] F. Piedad, High availability : design, techniques, and processes. Upper Saddle River: Prentice Hall, 2001.

[II-II-7] O. Yi, "Highly Available Database Systems," in Integrated information systems and databases seminar, 2006.

[II-8] C. P. Subbe, M. Kruger, P. Rutherford, and L. Gemmel, "Validation of a modified Early Warning Score in medical admissions," QJM, vol. 94, no. 10, pp. 521-526, 2001. [Online]. Available: http://qjmed.oxfordjournals.org/content/94/10/521.abstract

[II-9] L. Mota, "Nursing information systems: a study on the relevance of information." MSc Thesis (in Portuguese), Universidade do Porto, 2010.

[II-10] E. Ammenwerth, J. Brender, P. Nykanen, H.-u. Prokosch, M. Rigby, and J. Talmon, "Visions and strategies to improve evaluation of health information systems Reflections and lessons based on the HIS-EVAL workshop in Innsbruck," International Journal of Medical Informatics, 2004.

[II-11] S. Lameirao, "Hospital management and use of information systems: application to CHVR-PR," MSc Thesis (in portuguese), Universidade de Tras-os-Montes e Alto Douro, 2007.

[II-12] J. Machado, V. Alves, A. Abelha, and J. Neves,"Ambient intelligence via multiagent systems in the medical arena," Engineering Intelligent Systems for

Electrical Engineering and Communications, vol. 15, no. 3, pp. 151-157, 2007.

[II-13] J. Machado, A. Abelha, P. Novais, J. Neves, and J. a. Neves, "Quality of service in healthcare units," Int. J. Computer Aided Engineering and Technology, vol. 2, no. 4, pp. 436-439, 2010.

[II-14] M. P. Sullivan, "System Support for Software Fault Tolerance in Highly Available Database Management Systems," Tech. Rep., 1992.

[II-15] M. Cyran, P. Lane, and J. Polk, Oracle Database Concepts, 10g Release 2 (10.2). Oracle, 2005. [Online]. Available: http://docs.oracle.com/cd/B19306 01/server.102/b14220.pdf

[II-16] M. Bauer and R. Strohm, Oracle Real Application Clusters Administration and Deployment Guide, 11g Release 1 (11.1). Oracle, 2009. [Online]. Available: http://docs.oracle.com/cd/B28359 01/rac.111/b28254.pdf

[II-17] A. Albino and V. Jacinto, "Implementation of early warning score - MEWS," Centro Hospitalar do Barlavento Algarvio, EPE, Portimão (in portuguese), Tech. Rep., 2009.

[II-18] J. Gardner-Thorpe, N. Love, J. Wrightson, S. Walsh, and N. Keeling, "The value of Modified Early Warning Score (MEWS) in surgical inpatients: a prospective observational study." Annals of the Royal College of Surgeons of England, vol. 88, no. 6, pp. 571-5, Oct. 2006.

[II-19] C. P. Subbe, R. G. Davies, E. Williams, P. Rutherford, and L. Gemmell, "Effect of introducing the Modified Early Warning score on clinical outcomes, cardio-pulmonary arrests and intensive care utilisation in acute medical admissions," Anaesthesia, vol. 58, no. 8, pp. 797-802, 2003. [Online]. Available: http://dx.doi.org/10.1046/j.1365-2044.2003.03258.x

[II-20] G. Devaney and W. Lead, "Guideline for the use of the modified early warning score (MEWS)," Outer North East London Community Services, Tech.

Rep., 2011.

[II-21] S.Nair,“TheArtofDatabaseMonitoring” InformationSystemsControl Journal, no. Ccm, pp. 1-4, 2008.

[II-22] I. Chan, Oracle Database 2 Day + Performance Tuning Guide, 10g Release 2 (10.2). Oracle, 2010.

[II-23] K. Rich, Oracle Database Reference, 10g Release 2 (10.2). Oracle, 2009.

[II-24] I. Chan, Oracle Database Performance Tuning Guide, 10g Release 2 (10.2). Oracle, 2008.

[II-25] C. Shallahamer, Forecasting Oracle Performance. Berkeley: Apress, 2007.

[II-26] M. Ghazanfari, M. Jafari, and S. Rouhani, “A tool to evaluate the business intelligence of enterprise systems,” Scientia Iranica, no. 0, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S102630981100215X

[II-27] A. Nylund, “Tracing the BI Family Tree,” Knowledge Management, 1999.

[II-28] H. J. Watson and B. H. Wixom, “The Current State of Business Intelligence,” Computer, vol. 40, no. 9, pp. 96?99, 2007.

[II-29] S.-T. Li, L.-Y. Shue, and S.-F. Lee, “Business intelligence approach to supporting strategy-making of ISP service management,” Expert Syst. Appl., vol. 35, no. 3, pp. 739-754, 2008. [Online]. Available: http://dl.acm.org/citation.cfm?id=1383655.138375

[II-30] M. Tereso and J. Bernardino, “Open source business intelligence tools for SMEs,” in Information Systems and Technologies (CISTI), 2011 6th Iberian Conference on, 2011, pp. 1-4.

[II-31] C. Thomsen and T. Pedersen, “A Survey of Open Source Tools for Business Intelligence," in Data Warehousing and Knowledge Discovery, ser. Lecture Notes in Computer Science, A. Tjoa and J. Trujillo, Eds. Springer Berlin / Heidelberg, 2005, vol. 3589, pp. 74-84. [Online]. Available: http://dx.doi.org/10.1007/11546849 8

[II-32] Pentaho, "Welcome to the Pentaho Community." [Online]. Available: http://community.pentaho.com

[II-33] L. Ramos, "Performance Analysis of a Database Caching System In a Grid Environment," MSc Thesis (in portuguese), FEUP, 2007.

[II-34] K. Dias, M. Ramacher, U. Shaft, V. Venkataramani, and G. Wood, "Automatic performance diagnosis and tuning in Oracle," in Proceedings of the 2005 CIDR Conf, 2005. [Online]. Available: http://crmondemand.oracle.com/technetwork/database/focu areas/manageability/cidr-addm-134903.pdf

[II-35] J. Beresniewicz, "Average active sessions: the magic metric," Oracle USA, Tech. Rep.

[II-36] Hoopoes, "Oracle Database Tuning Statistics," 2007. [Online]. Available: http://www.hoopoes.com/cs/oracle tune.shtml

3.3 III - Intelligent Systems based in Hospital Database Malfunction Scenarios

Paulo Silva, César Quintas, Pedro Gonçalves, Gabriel Pontes, Manuel Santos, António Abelha and José Machado

4ht IEEE International Conference on Industrial Engineering and Engineering Management, Hong Kong, 2012.

3.3.1 Abstract

Databases are indispensable for everyday tasks in many organizations, particularly in healthcare units. Databases, allows to archive, among other relevant operations, important, private and confidential information about patients clinical status. Therefore, they must be available, reliable and at high performance level twenty-four hours a day, seven days per week. In many healthcare units, fault tolerant systems are online and ensure the availability, reliability and disaster recovery of data. However, these mechanisms do not allow to take preventive actions in order to avoid fault occurrence. In this context, it is of utmost importance the necessity of developing a fault prevention system. This system can predict database malfunction in advance and provides early decision taken to solve problems. With this paper we inted to monitor the database performance and adapt a forecasting model used in medicine (MEWS) to the database context. Based on mathematical tools it was created a scale that assesses the severity of abnormal situations. In this way, it is possible to define the scenarios where database symptoms must trigger alerts and assistance request.

3.3.2 Introduction

A database is composed by a set of hardware and software components and is coordinated by the database management system (DBMS). This system is also responsible for managing the database users' information requests [III-1][III-2]. In healthcare units it is crucial that database and DMBS ensure: data confidentiality, data integrity and data availability [III-2][III-3][III-4].

This paper focuses on the issue of availability of data. In healthcare units databases store very important information about the patients' clinical status, administrative information and other relevant information for the healthcare services. As so this data should be available twenty-four hours a day, seven days per week

[III-5]. However, ensuring the continuing availability of databases is not easy [III-6]. Faults often occur in the database. These faults can be of various types depending on the system where database are implemented. The most common type of faults in large databases under high utilization is related to resources limitation or high load. These faults occur when the physical resources are not enough to keep the database running. It is very important that the database must continue to operate despite the faults. Healthcare units have already implemented some fault tolerant systems, which intend to ensure the availability of data when a fault occurs. However, these systems only work after a fault occurs, not allowing to take preventive actions [III-7][III-8][III-9].

For that reason, the development of a fault forecasting and prevention system which allows taking early actions to solve future problems is crucial. Forecasting models have been used in critical areas such as medicine. The objective of this research is the development of a model for forecasting and prevention of database faults by adapting an existing forecasting model in medicine (Modified Early Warning Score - MEWS) [III-10].

This paper will present the information systems, the MEWS model, the methodology followed, the results obtained and finally the conclusions.

3.3.3 Healthcare Information Systems

The healthcare information systems allowed that, healthcare units, implement new methodologies for problem solving. These systems are responsible for the collection, storage, processing and data management of all stakeholders and all services. Thus, healthcare information systems enable a fast and efficient access to stored information always assuring data security [III-11].

Since the 1990s, hospital information systems in Portugal was implemented the SONHO (Hospital Information Integrated System). This system had as its objective merging clinical and administrative information. The evolution of this system is the genesis of the electronic medical record (EMR). An electronic medical record is a set of electronic documents that contain all information about patient health and his clinical risk profile [III-12] [III-13]. Besides the system already mentioned, in the healthcare units there are many other distributed and heterogeneous information systems. Therefore it is very difficult to ensure communication between the different systems. It became thus necessary to develop an effort to allow interoperability among all systems [III-13].

Therefore, a dynamic framework (AIDA- Agency for the Integration, Diffusion and Archive) was developed. This framework is composed of pro-active agents who are responsible for promoting communication. It made the management of all data most relevant to the process of interoperability between different systems. Thus, AIDA shares information and knowledge among every information system, namely: the administrative information system (AIS); the medical support information system (MIS); the nursing support information system (NIS); the EMR information system; the department information systems (DIS) of all the departments or of the healthcare unit [III-13][III-14].

The information is archived in very large databases which must be always available, because their information is vital for solving the patients' problems and for hospital management [III-5]. To achieve this objective fault tolerance mechanisms are used. These mechanisms can be based on the redundancy of data, components or both. They also can be used for load balancing or recovery [III-2][III-5].

The main databases (AIDA and SONHO) of Centro Hospitalar do Porto are based on an Oracle Real Application Clusters (RAC) System. This mechanism is provided by Oracle for improving the availability and scalability of databases. An example of these architectures can be see in Figure 3.13. These architectures are composed by a shared database which can be accessed through the server/computer that contains a database instance and an ASM (Automatic Storage Management) instance. So, it is possible access to the database across multiple servers. [III-2][III-8][III-15].

In AIDA database, there is also a data guard solution (Figure 3.14). A data guard solution consists in one or more standby databases (replicas of the original database) which should be in different places. Thus, when the master database is unavailable the replica can be used without the need to interrupt operation of the system. It is essential that the master and the standby databases are synchronized and the access is read-only during recovering [III-5].

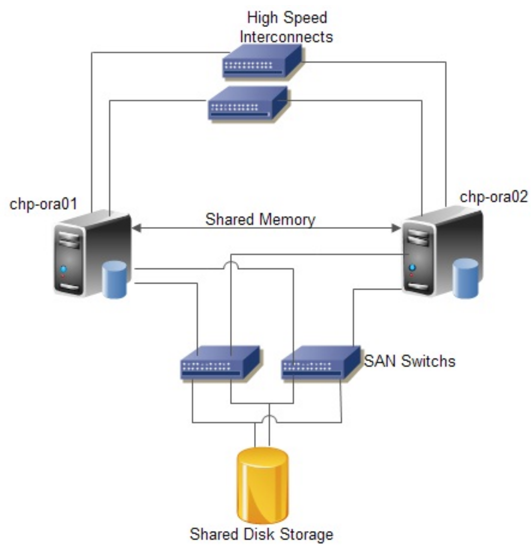

Figura 3.13: SONHO RAC Architecture

AIDA and SONHO are very large databases. AIDA includes around 5451 tables, a large amount of sessions at same time and several agents responsible for ensuring interoperability. SONHO has around 2255 tables and a large number of sessions and processes that run permanently. Therefore, the most often faults in these databases are related to resource limitation. Then the focus of this paper is to implement a forecasting model thact predict faults associated with resource overhead.

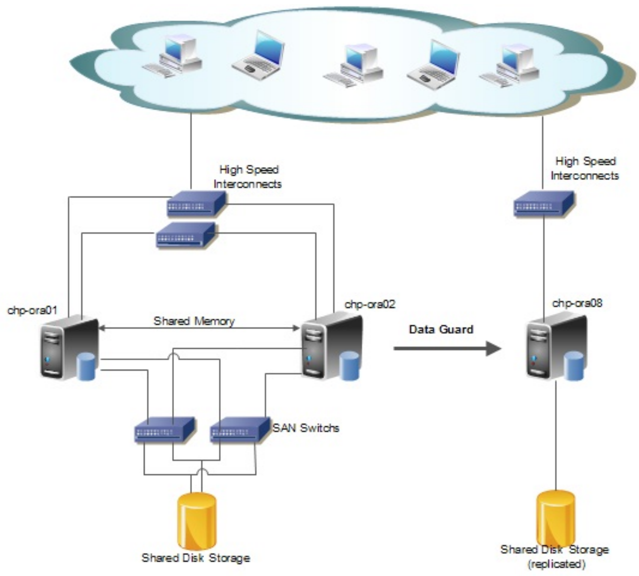

Figura 3.14: AIDA Complete Architecture

3.3.4 MEWS

In medicine there is already used a model, the Modified Early Warning Score(MEWS), for the prediction, in advance, of serious health problems. This model assumes that a serious problem of health is often preceded by physiological deterioration. It uses a decision table, like the Table 3.3 , to evaluate the clinical status of the patient according the monitoring patients' vital signs values [III-10][III-16][III-17].

Tabela 3.3: MEWS Score

Score	**3**	**2**	**1**	**0**	**1**	**2**	**3**
Temp (°C)		<35	35.1-36.0	36.1-38.0	38.1-38.5	>38.6	
HR (min -1)		<40	41-50	51-100	101-110	111-130	>130
Systolic (mmHg)	<70	71-80	81-100	101-199		> 200	
RR (min -1)		<8		8-14	15-20	21-29	>30
SPO2	<85	85-89	90-93	>94			
Urine (ml/Kg/h)	Nil	< 0.5					
Neurological				A	V	P	U

The Early Warning Score was been introduced in 1997 and the vital signs measured were: Temperature, Respiratory rate, Systolic BP, Heart rate and Neurological. However, some years later several studies have concluded that it was needed more parameters for the success of this model. In 1999 a new model appeared with the use of two additional variants: Urine output, Saturation of Hemoglobin with Oxygen SPO2 [III-16][III-17].

Normally, if any of the parameters have the score equal to two, the patient must be in observation. In the case of the sum of scores being equal to four, or there is an increase of two values in consecutive monitoring, the patient requires urgent medical attention. In a more extreme situation, if a patient has a score bigger than four, he is at risk of life [III-10][III-18].

The main advantage of this model is that it is possible identify the patients at risk and give them priority to access to the limited resources of intensive care units [III-16].

3.3.5 Methodology

A. Monitoring database performance

The first step to build a predictive model of faults is to try to identify the source of faults. Therefore, it is necessary to monitor the database performance [III-19].

Monitoring is not an easy process, and its complexity increases when it is necessary to perform the monitoring of various components and systems with complex architectures. However, the Oracle systems provide several tools to help in this like performance views. These views contain useful information for monitoring. They show the current state of the database, so it is impossible to realize editing operations, it is only possible to consult information [III-20][III-21].

The most important views related to this paper are the views which contain information about database component statistics such as v$sysstat or v$sesstat [III-20].

It is also necessary to use others tools more related with physical resources. There are a lot of operative system commands that can help in the monitoring process but the most used are **Top**, **Sar**, **Ps**, **Vmstat**. Using these tools together we can do a more efficient database monitoring [III-22].

B. Business Intelligence(BI) tool

In this paper, the use of BI tools is useful because database monitoring provides a large quantity of data which are difficult to interpret and manipulate and also because the BI tools allow you to perform useful for dashboards show real-time forecasts.

The term Business Intelligence (BI) was introduced by Howard Dresner of the Gartner Group, to describe a set of concepts, methods, processes which utilize new technologies to improve the business, particularly, on decision making. BI tools it is also important for cost-cutting in data, time savings for user requests and more and better information [III-23][III-24].

After reviewing some studies about open source BI tools and realize experiments involving some of these tools, it was observed that Pentaho Community version is the most appropriate tool to use in this work [III-25][III-26].

The Community version of Pentaho provides several tools that are available for different computational platforms. These tools allow the creation of reports, dashboards and charts, application of data mining and integration techniques and data modeling [III-25][III-27]. There is a main application, bi-server, where it is possible to perform reporting and analysis. In this application, can be added a others plugs-in such as CDE that allows the development of dashboards. In this paper, the tools most frequently used were [III-27]:

- **CDE** - Allows the user to develop dashboards, in order to show more clearly the data collected/treated. It was developed and maintained by Webdetails. The front-end are based in HTML, and dashboards can be populated with data coming from a variety of sources, such: sql querys, xml files, mondrian cubes, spoon transformations [III-27].

- **Pentaho Data Integration (Spoon)** - Delivers a powerful Extraction, Transformation and Loading (ETL) capabilities. This tool is very useful to integrate information from different sources and also to perform some mathematical operations on data [III-27].

C. Explain of the statistics

There are several statistics that can be used to characterize the behavior of the database. Since the objective is to prevent fault in terms the resource limitation have been selected statistics related to the load from the database. These are [III-5][III-20]:

- **DB time** - This statistic give information about database response time. The response time is the period between an initial user requests until the return of the results. It is a good indicator of the workload of the system. Typically, this time increases with the number of simultaneous users or applications, but also may increase due to large transactions, or other system problems [III-5][III-28].

- **Number of transactions** - Transactions is units of work, i.e., the database have more or less work according with transactions which are realize. In Oracle databases, the number of transactions can be obtained by summing the values of statistics "user commits" and "user rollbacks" [III-5][III-6][III-22].

- **Number of executions** - One transaction can be result in a high set of operations in database depending of query. Therefore, it is important to collect information about the number of operations [III-22].

- **Calls ratio** - Calls can be of two types: user calls or recursive calls. The user call occurs when a user request can be resolved through a single SQL query. A recursive call occurs when a user request need more than one SQL query. Ideally this ratio should be as low as possible, since the high number of recursive calls can indicate problems with the design of tables. This ratio can be calculate by the equation [III-21][III-29]:

$$RC = \frac{(recursivecalls)}{(recursivecalls + usercalls)} \tag{3.2}$$

- **Number of current logons** - Logons not directly represents the number of users but the number of sessions, since a user may have multiple sessions. The collection of this statistic is important because each session is associated with a piece of memory, so many simultaneous sessions can cause problems [III-20].

- **Processor utilization** - The processor is one of the most important statistic. Low values of processor utilization may indicate problems at the level of I/O. If the values are too high can compromise the functioning of the database [III-20].

- **Memory utilization** - The memory is a key component to the speed of the database systems, since, depending on whether the data are or are not in memory, the response time is influenced [III-6].

- **Size of redo file** - The redo files are used to store information about changes in the database. An increase in the size of these files, it indicates a higher number of operations and therefore a higher database load [III-21].

- **Buffer cache ratio** - This ratio show the percentage of data that are in memory cache, rather than disk. Normally, the BC is very high if BC decrease, this may indicate problems. BC can be calculate:

$$BC = \frac{1 - (physicalreads)}{(consistentegets + blockgets)} \tag{3.3}$$

- **Amount of I/O requests** - The I/O operations spend a lot of time, and an excess of this kind of operations can indicate problems in memory. Due problems of the memory the access to disc is more frequent and more time is spent in I/O operations [III-20].

- **Amount of redo space requests** - Indicates the lack of space to write in the buffer, this can lead to delays because it is necessary to write some data to disk to release memory. This can happen due to a poorly sized buffer, or excess entries generated simultaneously [III-29].

- **Volume of network traffic** - The several database components are connected through network. If a volume of network is increasing greatly, the database can be slow and compromise users' requests [III-20].

D. Process

First, a java application was developed to collect data related with statistics described above from the two major databases (AIDA, SONHO) of Centro Hopitalar do Porto. For a week the monitoring program collected the values of mentioned statistics. With these values it is possible to create charts, in Pentaho, that demonstrate normal behavior of the database. For each statistic, we calculated the upper (using the 75th percentile) and lower (using the 25th percentile) limits. Thus, it is possible to detect abnormal situations as those statistics on the measured value is not within limits.

A new program was developed for catching the abnormal situations. Depending on the value of the deviation and the frequency of times that happen abnormal situations are assigned granted scores like as is done in MEWS.

According to these scores, alerts will be sent to database administrator and it is possible visible in dashboard. This program is upgradeable as new limits calculated at the end of the day based on new measurements.

3.3.6 Results

The program is based in learning rules. Dynamically it creates a new model from normal database behavior. Figure 3.15 shows an example of a normal database (SONHO) behavior analyzing the statistic number of transactions per second.

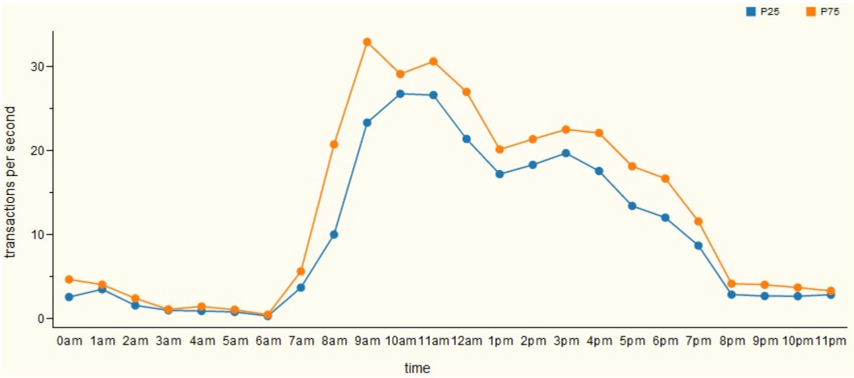

Figura 3.15: Transactions per second in SONHO

It is possible to see that the database has more transactions activity in the morning. Moreover, it is expected that 50% of the collect data are present between the two limits defined by the percentiles 25 and 75.

In order to create evaluation scores to identify the abnormal situations, it was calculated the percentiles 75, 80 and 90. The data above the percentile 75 are considered a low gravity situation, above 80 a grave situation and above 90 a critical situation. So, it is possible construct the decision table like in the Table 3.4.

Tabela 3.4: Severity Score

Score	**0**	**1**	**2**	**3**
Value	$< p75$	$< p75 < p80$	$< p80 < p90$	$> p90$
Severity	Normal	Low severity	Severe	Critical

The limits of the transactions per second can be seen in Figure 3.16.

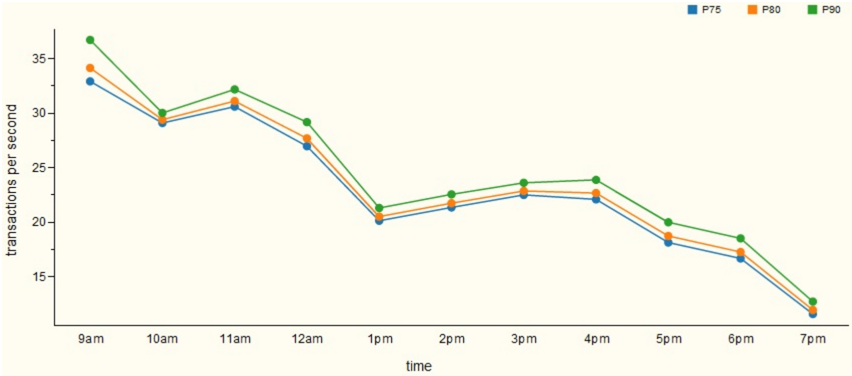

Figura 3.16: Percentiles 75, 80 and 90

For this test only are consider the period between 9am and 7pm because in this range there is the highest number of transactions per second.

Finally, the Figure 3.17 presents abnormal situations that occur along one day. According the previously scores, there is a low gravity situation at 9am and three critical situations at 11am, 12am and 1pm. In spite of these situations, the database continues in operation.

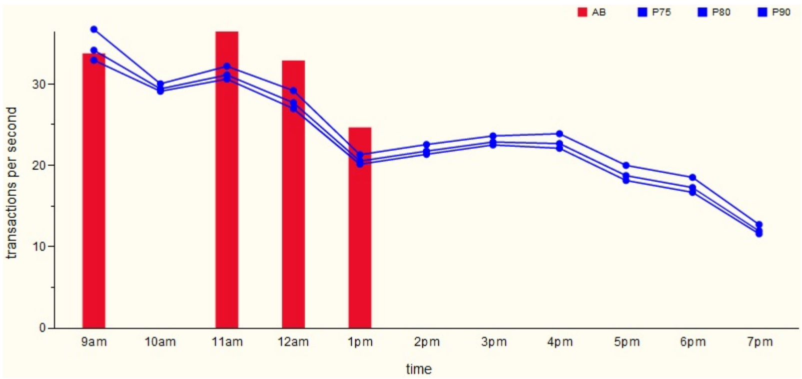

Figura 3.17: Abnormal situations: Transactions per second in SONHO

3.3.7 Conclusions

Despite fault forecasting in healthcare databases is a complex theme, it was proposed a model to represent the normal workload for databases AIDA and SONHO, based some chosen statistics. It was also possible to identify peak usage of databases and periods of the day where the possibility of faults occurrence is bigger.

With some mathematical tools it was possible to compute the limits for severity scores to characterize the abnormal situations. A decision table was created for this purpose. Thus, MEWS was adapted to the database context based in an upgradeable and learning forecasting fault system. The Pentaho Community tools were useful for manipulating data and for the development of the attractive dashboard for representing and interpretating results.

3.3.8 References

[III-1] T. Connolly and C. Begg, Database Systems A Pratical Approach to Design, Implementation, and Management, 4th ed. Harlow: Addison- Wesley, 2005.

[III-2] S. Drake, W. Hu, D. McInnis, M. Sko?ld, A. Srivastava, L. Thalmann, M. Tikkanen, O. y. Torbj rnsen, and A. Wolski, "Architecture of Highly Available Databases," in Service Availability, ser. Lecture Notes in Computer Science, M. Malek, M. Reitenspie, and J. Kaiser, Eds. Springer Berlin / Heidelberg, 2005, vol. 3335, pp. 1-16. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-30225-4 1

[III-3] A. Rodrigues, Oracle 10g e 9i: Fundamentos Para Profissionais. Lisboa: FCA, 2005.

[III-4] E. Bertino and R. Sandhu, "Database security - concepts, approaches, and challenges," Dependable and Secure Computing, IEEE Transactions on, vol. 2, no. 1, pp. 2-19, 2005.

[III-5] R. Godinho, "Availability, Reliability and Scalability in Database Architecture," MSc Thesis, Universidade do Minho, 2011.

[III-6] R. Schumacher, Oracle Performance Troubleshooting With Dictionary Internals SQL & Tuning Scripts. Kittrell: Rampant TechPress, 2003.

[III-7] O. Yi, "Highly Available Database Systems," in Integrated information systems and databases seminar, 2006.

[III-8] M. Cyran, P. Lane, and J. Polk, Oracle Database Concepts, 10g Release 2 (10.2). Oracle, 2005. [Online]. Available: http://docs.oracle.com/cd/B19306 01/server.102/b14220.pdf

[III-9] W. Hodak, S. Kumar, and A. Ray, "Oracle Database 11g High Availability," Oracle, Tech. Rep., 2007.

[III-10] C. P. Subbe, M. Kruger, P. Rutherford, and L. Gemmel, "Validation of a modified Early Warning Score in medical admissions," QJM, vol. 94, no. 10, pp. 521-526, 2001. [Online]. Available: http://qjmed.oxfordjournals.org/content/94/10/521.abstract

[III-11] E. Ammenwerth, J. Brender, P. Nykänen, H.-u. Prokosch, M. Rigby, and J. Talmon, "Visions and strategies to improve evaluation of health information systems Reflections and lessons based on the HIS-EVAL workshop in Innsbruck," International Journal of Medical Informatics, 2004.

[III-12] S. Lameirão, "Hospital management and use of information systems: application to CHVR-PR," MSc Thesis (in portuguese), Universidade de Trás-os-Montes e Alto Douro, 2007.

[III-13] J. Machado, V. Alves, A. Abelha, and J. Neves, "Ambient intelligence via multiagent systems in the medical arena," Engineering Intelligent Systems for Electrical Engineering and Communications, vol. 15, no. 3, pp. 151-157, 2007.

[III-14] J. Machado, A. Abelha, P. Novais, J. Neves, and J. a. Neves, "Quality of service in healthcare units," Int. J. Computer Aided Engineering and Technology, vol. 2, no. 4, pp. 436-439, 2010.

[III-15] M. Bauer and R. Strohm, Oracle Real Application Clusters Administration and Deployment Guide, 11g Release 1 (11.1). Oracle, 2009. [Online]. Available: http://docs.oracle.com/cd/B28359 01/rac.111/b28254.pdf

[III-16] A. Albino and V. Jacinto, "Implementation of early warning score - EWS," Centro Hospitalar do Barlavento Algarvio, EPE, Portimão (in portuguese), Tech. Rep., 2009.

[III-17] J. Gardner-Thorpe, N. Love, J. Wrightson, S. Walsh, and N. Keeling, "The value of Modified Early Warning Score (MEWS) in surgical inpatients: a prospective observational study." Annals of the Royal College of Surgeons of England, vol. 88, no. 6, pp. 571-5, Oct. 2006.

[III-18] G. Devaney and W. Lead, "Guideline for the use of the modified early warning score (MEWS)," Outer North East London Community Services, Tech. Rep., 2011.

[III-19] S.Nair,"TheArtofDatabaseMonitoring,"InformationSystemsControl Journal, no. Ccm, pp. 1-4, 2008.

[III-20] I. Chan, Oracle Database 2 Day + Performance Tuning Guide, 10g Release 2 (10.2). Oracle, 2010.

[III-21] K. Rich, Oracle Database Reference, 10g Release 2 (10.2). Oracle, 2009.

[III-22] C. Shallahamer, Forecasting Oracle Performance. Berkeley: Apress, 2007.

[III-23] M. Ghazanfari, M. Jafari, and S. Rouhani, "A tool to evaluate the business intelligence of enterprise systems," Scientia Iranica, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S102630981100215X

[III-24] A. Nylund, "Tracing the BI Family Tree," Knowledge Management, 1999.

[III-25] M. Tereso and J. Bernardino, "Open source business intelligence tools for SMEs," in Information Systems and Technologies (CISTI), 2011 6th Iberian

Conference on, 2011, pp. 1-4.

[III-26] C. Thomsen and T. Pedersen, "A Survey of Open Source Tools for Business Intelligence," in Data Warehousing and Knowledge Discovery, ser. Lecture Notes in Computer Science, A. Tjoa and J. Trujillo, Eds. Springer Berlin / Heidelberg, 2005, vol. 3589, pp. 74-84. [Online]. Available: http://dx.doi.org/10.1007/11546849 8

[III-27] Pentaho, "Welcome to the Pentaho Community." [Online]. Available: http://community.pentaho.com/

[III-28] J. Beresniewicz, "Average active sessions: the magic metric," Oracle USA, Tech. Rep.

[III-29] Hoopoes, "Oracle Database Tuning Statistics," 2007. [Online]. Available: http://www.hoopoes.com/cs/oracle tune.shtml

3.4 IV - Intelligent and Real Time Data Acquisition and Evaluation to Determine Critical Events in Intensive Medicine

Filipe Portela, Pedro Gago, Manuel Filipe Santos, José Machado, António Abelha, Álvaro Silva, Fernando Rua, César Quintas e Filipe Pinto

HCIST 2012 - International Conference on Health and Social Care Information Systems and Technologies

3.4.1 Abstract

Using the information regarding critical events to support decision making in Intensive Care Units would be useful. However it is seldom used in real settings as the information regarding those critical events is difficult to gather and make available in real time. The most usual procedures record only those events that are related to errors. This paper presents a solution to obtain critical events from clinical data. From data collected using an automatic and real-time data acquisition system it is possible to calculate the critical events regarding five variables that are usually monitored in an ICU. These results are presented to the medical and nursing staff in a friendly and intuitive mode. Using a color code our system provides visual warnings related to the evolution of the monitored variables values. Actually, a quick glance allows doctors to get a high level overview of the patient's condition.

3.4.2 Introduction

Intensive Care Units (ICU) are flooded with devices aiming at giving doctors better information regarding their patient's condition. For example, bed side monitors continuously show blood pressure, oxygen saturation, heart rate together with the values of other variables.

When working with such data we are faced with a wealth of data streams instead of the more usual static data table [IV-1]. Indeed, for each monitored variable in the ICU one has a data stream and, to further complicate matters, it is often the case that data from different sensors are gathered at different time intervals. For example, real time data, acquired from the bedside monitors must be merged together with analysis data, which comes in only once or twice a day. Thus, while traditionally one would have one or two records per patient per day, entered manually by the nursing and/or medical staff, automated data acquisition gives us an almost infinite amount of data. In fact, in the ICU we can now collect data for 16 different variables (when applied) with a 10 second interval between observations. This scenario makes us face a potentially high number of information per patient, per day. And, as already stated, this data arrives at different times and with different time intervals.

For automatic data acquisition to be possible and efficient in such a scenario, data pre-processing is of paramount importance assuring that important information is kept and that irrelevant one is filtered out. For example, one does not need to store the same amount of data for a stable patient that is needed for another that is having more problems. Also, doctors seldom need the exact values, they would rather reason in terms of change (e.g. "did the blood pressure rise much in the last hour?") than in terms of the exact values each variable had.

One very important pre-processing activity in such a context is that of identifying those situations that are potentially medically relevant as there is a huge number of value changes and only a small subset is potentially medically relevant. Models used for outcome and organ failure prediction depend on information regarding Critical Events (CE). In order to use those models in a real setting it was necessary to define the procedures to automatically compute CE for five variables: Diuresis, Blood Pressure, Heart Rate, Respiratory and Temperature.

This paper is divided in seven chapters, the first (this one) introduced the paper, the second frames the problem, then the data acquisition process and data analysis will be described. Chapter five will present the CE tracking system and the sixth will present the results obtained so far and the system interface. Finally some concluding considerations will be made.

3.4.3 Background

A. INTCare

The work we are presenting is part of a research project called INTCare. The algorithms in use in INTCare's prediction module require information regarding the critical events that have occurred during the stay of the patient in the ICU. In order to be able to use INTCare with real time data one must have some means of extracting the CE also in real time.

INTCare includes an Electronic Nursing Record (ENR) module that is responsible for data collection [IV-2]. It maintains the patients daily clinical records and other information relevant to the decision making process. Data collected includes the patient's vital signs, medical procedures, therapeutic plans, medical scores, information regarding ventilation and others. There is a touch screen workstation near each ICU bed. Medical and nursing staff may use them to record, validate and visualize all clinical information about each patient.

INTCare [IV-3] [IV-4] is implemented in the Intensive Care Unit of Hospital Santo Antnio, Centro Hospitalar do Porto.

B. Intensive Care

Intensive care is a critical area of medicine, where the patients are in too weak conditions and/or in serious life-risk [IV-5]. Intensive Care units are the place where this type of medicine is applied. Usually, during the stay of these patients is possible to verify a set of adverse events. These events can influence the future outcome and can occur various times a day [IV-6]. The ability to calculate the events automatically and in real-time is an important support to the decision making process. In addition, the development of an Intelligent Decision Support System, like INTCare, which uses these variables to predict the patient condition, can help the doctors to maintain a pro-active action that will benefits the patients.

C. Critical Events

Studies done in the past reported that the most common adverse errors were due to wrong mechanical or human performance [IV-7]. However, other points exist, that are difficult to analyze (eg. patient clinical events). This happens because it is very difficult to quantify the number of clinical errors due the lack of automatic data acquisition systems in the ICUs. Normally, these results are collected by some alerts provided from bedside monitors ([IV-8]). This paper will explain an approach to obtain the number and the time of clinical adverse events to five variables (Heart Rate, Blood Pressure, Saturation of Oxygen, Diuresis and Temperature).

The adverse events in use were assigned in an electronic application at a continuous acquisition basis. To understand if an event is critic or not, two main criteria were used [IV-9]:

- occurrence and duration should be registered by physiological changes;
- related physiological variables should be routinely registered at regular intervals.

An event is considered critical, when a longer event occurs or a more extreme physiologic measurement is found [IV-9]. In this project we used two different definitions: critical values and critical events. Critical values are values that are out of a normal range. Critical event is defined as a label to signal that a variable had critical values for more than the admissible time span, as defined in Table 3.5. Also, a critical event may signal that the critical value was so out of range that it is considered serious regardless of the duration of that observation. For example, a critical event happens whenever the patient's heart rate stays above 120 bpm for more than

1 hour. Also, a critical event happens every time the heart rate drops below 30 bpm or rises above 180 bpm.

Tabela 3.5: The protocol for the out of range physiologic measurements (adapted from [IV-9])

	BP	**Sp02**	**HR**	**UR**
Normal range	$90 to 180 mmHg$	$>= 90\%$	$60 to 120 bpm$	$>= 30 ml/h$
Critical event (a)	$>= 1h$	$>= 1h$	$>= 1h$	$>= 2h$
Critical event (b)	$< 60 mmHg$	$< 80\%$	$< 30 or > 180 bpm$	$<= 10 ml/h$

(a) Defined when continuously out of range

(b) Defined anytime

3.4.4 Data acquisition process

To implement the system it is necessary to be able to gather the values for all variables automatically, continuously and in real-time. Initially, none of the variables was acquired automatically or/and in an electronic mode. Every hour these values were registered by hand in a paper based nursing record. After some changes [IV-2] [IV-4] [IV-10] were made in the ICU acquisition system it became possible to automatically collect some data. Now, using these new data sources it is possible to make a calculation of the critical events. For the critical events system, the following data sources are being used:

- Bed Side Monitor - vital signs (VS);
- Electronic Nursing Record (ENR) ? Hourly values (Diuresis);

To perform the tasks associated with data acquisition a set of agents have been implemented [IV-11] to acquire and process the data from the different ICU data sources.

Firstly, the data are extracted from the ICU data sources by the agents; during the extraction some automatic procedures are executed. Then, all process of interpretation of values is executed by the pre-processing agent and finally, all data are loaded into a data warehouse that stores all critical values to be used by the Critical Events System. Figure 3.18 summarizes the process of the data acquisition system which has been developed in order to collect data from bedside monitors and ENR.

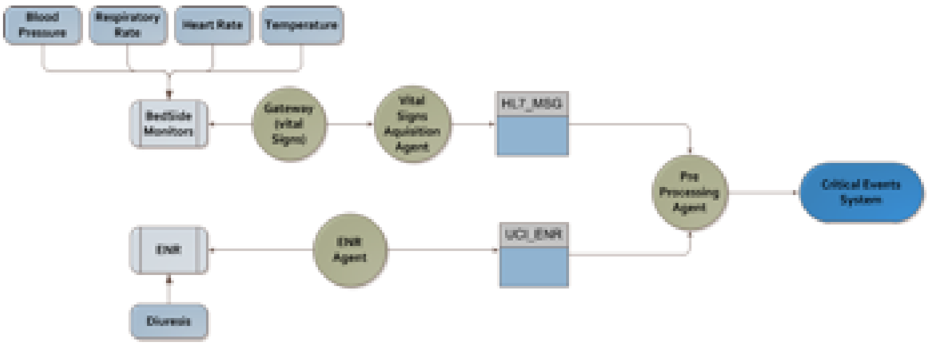

Figura 3.18: Transactions per second in SONHO

In the present we are concentrating on detecting the critical events that were deemed most important by the doctors we are working with. Automated data acquisition without human intervention is already in place for most of them. However, the urine output measurement is not automated at the ICU and the respective values have to be manually entered by the nursing staff. Table 3.6 explains, which are the variables collected in real-time, the correspondent data source and if they are acquired in an automatic or manual way. In this case, only one of the variables requires a manually operation.

Tabela 3.6: Critical variables data source and acquisition type

Variable	Data Source	Acquisition
Blood Pressure (BP)	BM	Automatic
Temperature (TEMP)	BM	Automatic
Heart Rate (HR)	BM	Automatic
Oxygen Saturation (02)	BM	Automatic
Urine Output (UR)	ENR	Manual (hour)

3.4.5 Data Analysis

This process is the main process of the entire project. It is a first step through the data that allows for validation and for detection of values that may be part of CE. The data quality is fundamental to obtain the correct number of critical events. In this case, all values collected are being used, i.e., all values are necessary to be able to do a continuous calculation of events. In order to streamline the process it is necessary to have an automatic validation procedure. This procedure is activated whenever a new value arrives. Each value is checked to see if it is inside the valid range of values for the respective variable and then it is flagged as potentially critical or non-critical. This task is executed according to the values defined in Table 3.7. For example, a value of 25 for the temperature will be discarded as it is out of the valid range (25C is the room temperature and is sometimes recorded as the patient's temperature if the sensor is not properly placed). A value of 39 for the same variable will be flagged as potentially critical.

Tabela 3.7: Data ranges

EVID	**DESCR**	**MINEC**	**MAXEC**	**MINVALICU**	**MAXVALICU**	**MINANYTIME**	**MAXANYTIME**
3510	TEMP	36	38	34	45	35	40
1011	BP	90	180	0	300	60	
3000	O2	90	100	0	100	80	
2009	HR	60	120	0	300	30	180
DIU	UR	30	1000	0	1000	10	

The procedure referred is applied only to the variables that are to be used in the calculation of the critical events. This procedure is executed through a trigger which is activated before the value is inserted in the table containing HL7 [IV-12] data. For each value collected:

```
BEGIN
  IF CATEGORY IN (3000, 2009, 1011, 3510, DIU ) THEN
    IF VALUE >= MINVALUEICU AND VALUE <= MAXVALUEICU THEN
      IF (VALUE <= MINANYTIME OR VALUE >= MAXANYTIME) THEN
```

```
        SET CRITIC TO 2
        SET VALID TO 1
      ELSIF VALUE <= MINEC OR VALUE >= MAXEC THEN
        SET CRITIC TO 1
        SET VALID TO 1
      ELSE
        SET CRITIC TO 0
        SET VALID TO 1
      ENDIF
    ENDIF
  ENDIF
END
```

Immediately after this procedure is executed another trigger, which is responsible to calculate the time of these events, is activated.

3.4.6 Critical Events Tracking System

After having collected the critical values it is necessary to check if they originated a critical event. That requires some calculations. The base of the calculation is the Table 1. In order to help understand if the event type collected is or is not the same which was before collected, a flag in the table will be used. The flag ϕ will identify the state of the event type collected, i.e., the flag value present in the table notifies if exist some event in the table with the same type and if that event is opened (1) or not (0). An "open" event is still in progress while a "closed" one has already finished.

When a value is collected and, immediately after it is validated and inserted in the HL7 table, the trigger will verify if there is some "open" record for this event category. If it doesn't exist, a new row will be created. If there is a record for the same event type, nothing will be done, otherwise the event finish date will be defined according the collected date of the obtained value.

For each event type (0, 1, 2) some operations will be done according to the respective event state (open or not). Finally and after the event start and event finish date is filled, a procedure, which calculates the time of each event is executed. For each patient, event category and value collected (PEVID) and, after the values are correctly inserted in database, the trigger shown next will be executed:

```
BEGIN
  IF OPEN = 0 THEN
    SET DATESTART TO SYSDATE
    SET DATEFINISH TO NULL
    SET OPEN TO 1
    SET CRITIC TO EVTYPE
  ENDIF
  IF OPEN = 1 THEN
    IF CRITIC <> EVTYPE THEN
      SET DATEFINISH TO SYSDATE
      SET OPEN TO 0
      SET CRITIC TO EVTYPE
      SET TOTALTIME TO DATESTART - DATEFINISH
    ELSE
      NULL
    ENDIF
  ENDIF
END
```

After the value is correctly identified and inserted into the database, with total time filled, another trigger will be executed. This other trigger is designed to verify if an event is critic or not.

To identify if a set of collected results is or not a critical event it is necessary to analyze the table which contains all values collected with the respective time interval. By computing the length of that time interval, the procedure will identify which set of values are critical:

```
BEGIN
  READ ROW
  IF EVTYPE = 2 THEN
    SET CRITICALEVENT TO 1
  ELSIF EVTYPE = 1
    IF CATEGORY = 'DIU'
      IF TOTALTIME >= 7200 THEN
        SET CRITICALEVENT TO 1
      ELSE
        SET CRITICALEVENT TO 0
      ENDIF
    ELSIF TOTALTIME >= 3600 THEN
      SET CRITICALEVENT TO 1
    ELSE
      SET CRITICALEVENT TO 0
    ENDIF
  ENDIF
END
```

3.4.7 Results

As results, it is possible obtain a number of critical events for the patient by hour and category. The system is refreshed every 10 minutes. The results are obtained after know the importance of the values, i.e. critic or not, other procedure will start. This procedure will calculate the Accumulated Critical Events (ACE)

- to reflect the patients' clinical evolution/severity of illness by hour. The values obtained will be used to create some ratios. The next procedures will be in a regular execution, and have the objective to understand the number of events by type and hour, and the sum of them.

To count the number of events and sum of the time, grouping by type and hour:

SET CATEGORYHOURVALUE TO SUM(COUNT(CRITICALEVENTS) BY DATEHOUR AND BY CATEGORY)
SET CATEGORYHOURTIME TO SUM(TOTALTIME) BY DATEHOUR AND BY CATEGORY)

The results obtained by the last procedures will be presented inside of Electronic Nursing Record in a grid designed to the effect. This system is composed by a traffic light system and a grid. The light system is an alert to the patient condition. For each variable, if some event is "open", the event type (1 or 2) is checked. Then, for the event type = 1, it will calculate the time between the system date (sysdate) and the starting date. If the time in minutes is between 10 and 20 the label will be yellow. If this event is open more than 20 minutes the label will be red. In the cases of event type = 2 the label will be always red independent the time it is started. In the other cases the label will be green. The critical event number and time results are presented in a grid. This grid shows: the number of critical events by hour, the number of accumulated events and the time in critical event by hour and the total time in critical event. The grid has 13 columns. The first represents the hours of the day (1-24), the next four columns, contain the number of events for that hour for each event. Then, the next four columns contain the number of minutes of the event in this hour. The last two columns represent the number total of events and the ratio minutes by hour for all events. In a context of proving the functionality of the system, Figure 3.19 present the distribution of the values collected during the last six months. The values are grouped by variable, event type and hour. In the graph is possible to see the percentage of occurrence for three of the five variables (blood pressure (BP), SPo2, Heart Rate (HR) by hour (0-23) and event type (1 or 2).

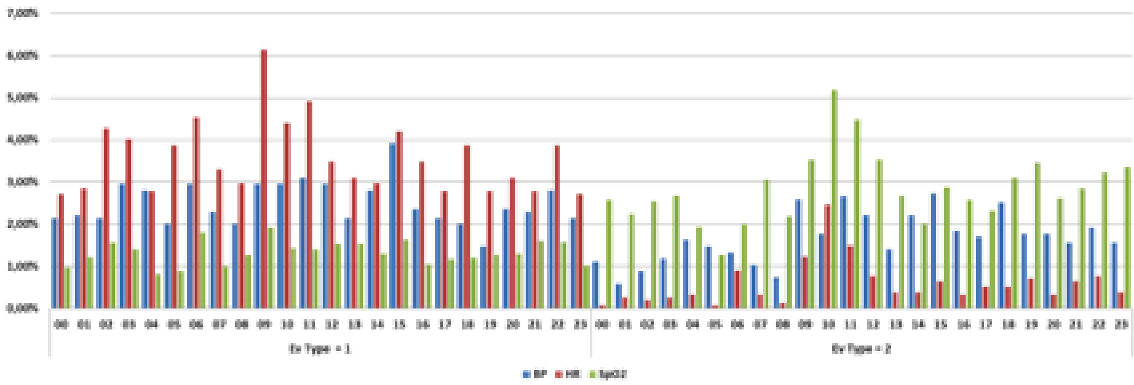

Figura 3.19: ICU Critical Events - Data acquisition

From the graph some important conclusions can be made. The critical events of BP (59,85%) and HR (85,91%) are strongly associated to the events type 1. In the case of SPo2 the critical events are mostly of the type 2 (68,22%). The vast majority of the events (almost 20%) occurs between 9 and 11 am. This is most evident for the HR (critical event type 1) and SpO2 (critical event type 2). For the Blood Pressure, the 15th hour is the most critical (6,64% = 3,91% (1) + 2,73% (2) of the critical events occurred during this hour). This sort of analysis can help the doctors understand when each the event type normally occurs. The information obtained about the critical events combined with other variables can be used to alert the doctors in order to avoid critical events.

3.4.8 Conclusions & Future Work

The implementation of this new approach allows doctors, in ICU of Centro Hospital do Porto, to have a better understanding of the patient's condition. The doctors can see, for each patient and in real time the number of events by hour and the time which the patient was in a critical event. In addition, they possess a traffic light system (green, yellow, red) to alert / show the actual situation for each event.

The output of the system developed will feed the DM models of the INTCare system in order to predict the organ failure and outcome. Due to the success of this implementation, in the future, further adverse events types will be explored and added to the system.

3.4.9 References

[IV-1] R. L. Mador and N. T. Shaw, "The impact of a Critical Care Information System (CCIS) on time spent charting and in direct patient care by staff in the ICU: a review of the literature," International Journal of Medical Informatics, vol. 78, pp. 435-445, 2009.

[IV-2] F. Portela, P. Gago, M. F. Santos, A. Silva, F. Rua, J. Machado, A. Abelha, and J. Neves, "Knowledge Discovery for Pervasive and Real-Time Intelligent Decision Support in Intensive Care Medicine," presented at the KMIS 2011- International Conference on Knowledge Management and Information Sharing., Paris , France, 2011.

[IV-3] F. Portela, M. F. Santos, and M. Vilas-Boas, "A Pervasive Approach to a Real-Time Intelligent Decision Support System in Intensive Medicine," in Communications in Computer and Information Science. vol. 0272, Springer, Ed., ed, 2012, p. 14.

[IV-4] F. Portela, M. F. Santos, P. Gago, Á. Silva, F. Rua, A. Abelha, J. Machado, and J. Neves, "Enabling real-time intelligent decision support in intensive care," 2011.

[IV-5] N. Bricon-Souf and C. R. Newman, "Context awareness in health care: A review," International Journal of Medical Informatics, vol. 76, pp. 2-12, 2007.

[IV-6] J. M. Rothschild, C. P. Landrigan, J. W. Cronin, R. Kaushal, S. W. Lockley, E. Burdick, P. H. Stone, C. M. Lilly, J. T. Katz, and C. A. Czeisler, "The Critical Care Safety Study: The incidence and nature of adverse events and serious medical errors in intensive care*," Critical care medicine, vol. 33, p. 1694, 2005.

[IV-7] M. Kaur, M. Pawar, J. K. Kohli, and S. Mishra, "Critical events in intensive care unit," Indian Journal of Critical Care Medicine: Peer-reviewed, Official Publication of Indian Society of Critical Care Medicine, vol. 12, p. 28, 2008.

[IV-8] M. T. Keegan, O. Gajic, and B. Afessa, "Severity of illness scoring systems in the intensive care unit," Critical care medicine, vol. 39, p. 163, 2011.

[IV-9] Á. Silva, P. Cortez, M. F. Santos, L. Gomes, and J. Neves, "Rating organ failure via adverse events using data mining in the intensive care unit," Artificial Intelligence in Medicine, vol. 43, pp. 179-193, 2008.

[IV-10] F. Portela, M. F. Santos, Á. Silva, J. Machado, and A. Abelha, "Enabling a Pervasive approach for Intelligent Decision Support in Critical Health Care," presented at the HCist 2011 ? International Workshop on Health and Social Care Information Systems and Technologies, Algarve, Portugal, 2011.

[IV-11] M. F. Santos, F. Portela, M. Vilas-Boas, J. Machado, A. Abelha, and J. Neves, "INTCARE - Multi-agent approach for real-time Intelligent Decision Support in Intensive Medicine," in 3rd International Conference on Agents and Artificial Intelligence (ICAART), Rome, Italy, 2011.

[IV-12] J. S. Hooda, E. Dogdu, and R. Sunderraman, "Health Level-7 compliant clinical patient records system," 2004, pp. 259-263.

Capítulo 4

Conclusão e Trabalho Futuro

4.1 Conclusão

O propósito deste trabalho centrou-se no estudo conducente a modelos de previsão de eventos críticos associados aos Sistemas de Gestão de Bases de Dados (SGBD), peças chave dos sistemas de informação, por forma a evitar situações de paragem. Por outro lado, o estudo procurou também investigar a possibilidade de utilizar modelos de previsão quer para necessidades futuras de armazenamento quer para os recursos computacionais necessários.

Os principais resultados deste trabalho, assim como a sua contribuição científica, emergem dos quatro estudos efetuados e que foram alvo de publicação. De seguida apresenta-se uma sinopse destes estudos.

- **Estudo I -** *Hospital database workload and fault forecasting*

  No âmbito deste estudo, foi proposto um modelo para caracterizar e representar a carga de trabalho normal para a base de dados da plataforma AIDA. Foram determinados os limites que caracterizam as situações anormais através de percentis. Foi definida uma tabela de decisão, com os percentis e scores recorrendo à metodologia MEWS (um modelo utilizado na medicina intensiva). A título de curiosidade, observou-se que o nodo 1 da base de dados é mais usado do que nodo 2. Além disso, foram identificados picos de carga de trabalho e foi possível observar que o período mais crítico é o período compreendido entre as 10:00 da manhã e as 00:00 da noite.

- **Estudo II -** *Step Towards Fault Forecasting in Hospital Information Systems*

  Este trabalho mostra o quão importante é o tema de previsão de falhas em bases de dados na área da saúde. Apesar de ser um tema complexo, foi proposto um modelo para representar a carga de trabalho normal para as bases de dados dos sistemas AIDA e SONHO. Com este modelo é possível fazer diversas análises e com a ajuda de outras ferramentas matemáticas é possível desenvolver um programa que detete as situações anormais. De acordo com a frequência e intensidade destas situações anormais é possível criar uma tabela de decisão com as contagens que representam a gravidade situação. As pontuações representam a gravidade da situação. A abordagem proposta pela metodologia MEWS foi adaptada à realidade das bases de dados através de um sistema de previsão de falhas adaptativo.

- **Estudo III -** *Intelligent Systems based in Hospital Database Malfunction Scenarios*

  Apesar de previsão de falhas em bases de dados de saúde se revestir de bastante complexidade, foi proposto um modelo para representar a carga de trabalho normal para as bases de dados dos sistemas AIDA e SONHO, com base algumas estatísticas selecionadas. Neste estudo são propostos limites para os scores de gravidade para caracterizar as situações anormais.

- **Estudo IV -** *Intelligent and Real Time Data Acquisition and Evaluation to Determine Critical Events in Intensive Medicine*

  A implementação desta nova abordagem permite que os médicos das unidades de cuidados intensivos do Centro Hospitalar do Porto passem a ter uma melhor compreensão da condição do paciente. Os médicos podem ver, para cada paciente e em tempo real, o número de eventos por hora e o tempo que o paciente estava em um evento crítico. Além disso, eles possuem um sistema de semáforo (verde, amarelo, vermelho) para alertar / mostrar a situação real de cada evento. A saída do sistema desenvolvido vai alimentar os modelos de data mining do sistema INTCare, a fim de prever a falha do órgão e o outcome. Devido ao sucesso desta aplicação, no futuro, outros tipos de eventos adversos serão explorados e adicionados ao sistema.

4.2 Trabalho Futuro

Os resultados obtidos permitem sustentar novos desenvolvimentos e novas investigações, levando a aplicação da metodologia MEWS para outros âmbitos. Como trabalho futuro, propõe-se:

- O refinamento dos modelos apresentados utilizando dados entretanto recolhido sobre o comportamento das bases de dados e dos modelos de previsão (adaptabilidade);
- A aplicação da abordagem baseada na metodologia MEWS a outras bases de dados.

Bibliografia

[Abelha, 2004] Abelha, A. (2004). *Multi-Agent Systems to Support Cooperative Work in Health Care Units*. PhD thesis, Universidade do Minho.

[Abelha et al., 2007] Abelha, A., Analide, C., Machado, J., Neves, J., Santos, M., and Novais, P. (2007). Ambient intelligence and simulation in health care virtual scenarios. In *Establishing the Foundation of Collaborative Networks. Camarinha-Matos, L., Afsarmanesh, H., Novais, P., Analide, C. (Eds)*, ISBN: 978-0-387-73797-3, pages 461–468. Springer-Verlag.

[Abelha et al., 2004] Abelha, A., Machado, J., Alves, V., and Neves, J. (2004). Data warehousing through multi-agent systems in the medical arena. In *Proceedings of the 1st International Conference on Knowledge Engineering and Decision Support*, Porto, Portugal.

[Abelha et al., 2002] Abelha, A., Machado, J., Santos, M., Allegro, S., Rua, F., Paiva, M., and Neves, J. (2002). Agency for integration, diffusion and archive of medical information. In *Proceedings of the Third IASTED International Conference - Artificial Intelligence and Applications*, ISBN: 0-88986-390-3, Malaga, Spain.

[Aier and Schönherr, 2006] Aier, S. and Schönherr, M. (2006). Evaluating integration architectures – a scenario based evaluation of integration technologies. *Trends in Enterprise Application Architecture*, pages 2–14.

[Berg, 2009] Berg, M. (2009). *Health Information Management: Integrating Information Technology in Health Care Work*. Routledge.

[Bertino and Sandhu, 2005] Bertino, E. and Sandhu, R. (2005). Database security - concepts, approaches, and challenges. *Dependable and Secure Computing, IEEE Transactions*, 2(1):2–19.

[Chan, 2008] Chan, I. (2008). *Oracle Database Performance Tuning Guide, 10g Release 2 (10.2)*. Oracle.

[Dick and Steen, 1991] Dick, R. and Steen, E. (1991). *The Computer-based Patient Record - An Essential Technology for Health Care*. National Academy Press, Washington, DC.

[Dolin et al., 2001] Dolin, R., Alschuler, L., Boyer, S., Beebe, C., Behlen, F., Biron, P., and Shvo, A. (2001). Hl7 clinical document architecture, release 2. *Journal of the American Medical Informatics Association*, 13(1):30–39.

[Dreiseitl and Binder, 2005] Dreiseitl, S. and Binder, M. (2005). Do physicians value decision support? a look at the effect of decision support systems on physician opinion. *Artificial Intelligence in Medicine*, 33(1):25–30.

[Duarte, 2008] Duarte, J. (2008). Qualidade e normalização do registo no processo clínico electrónico. Master's thesis, Universidade do Minho.

[Duarte, 2015] Duarte, J. (2015). *Improving Quality on Medical Records with Medical Onthologies*. PhD thesis, Universidade do Minho.

[Duarte et al., 2011] Duarte, J., Portela, F., Abelha, A., Machado, J., and Santos, M. (2011). Electronic health record in dermatology service. In *ENTERprise Information Systems - International Conference, CENTERIS 2011, Vilamoura, Portugal, October 5-7, 2011, Proceedings, Part III*, volume 221 of *Communications in Computer and Information Science*, pages 156–164. Springer Berlin Heidelberg.

[Dupuits and Hasman, 1995] Dupuits, F. and Hasman, A. (1995). Evaluation of a decision support system in a medical environment. *International Journal of Bio-Medical Computing*, 38(2):155–165.

[Erl, 2004] Erl, T. (2004). *Service-Oriented Architecture. A Field Guide to Integrating XML and Web Services.* Prentice hall.

[Finlay, 2004] Finlay, P. (2004). *Introducing decision support systems.* NCC Blackwell, 2nd edition edition.

[Gardner-Thorpe et al., 2006] Gardner-Thorpe, J., Love, N., Wrightson, J., Walsh, S., and Keeling, N. (2006). The value of modified early warning score (mews) in surgical in-patients: a prospective observational study. *Annals of The Royal College of Surgeons of England*, 88:571.

[Godinho, 2011] Godinho, R. (2011). *Availability, Reliability and Scalability in Database Architecture.* PhD thesis, Universidade do Minho.

[Haux, 2006a] Haux, R. (2006a). Health information systems - past, present, future. *International Journal of Medical Informatics*, 75:268–281.

[Haux, 2006b] Haux, R. (2006b). Individualization, globalization and health - about sustainable information technologies and the aim of medical informatics. *International Journal of Medical Informatics*, 75:795–808.

[Hodak et al., 2007] Hodak, W., Kumar, S., and Ray, A. (2007). Oracle database 11g high availability. Technical report, Oracle.

[Hong, 2006] Hong, X. (2006). A database monitoring and disaster recovery system. In *Frontiers of WWW Research and Development - APWeb 2006*, volume 2841, pages 1201–1204. Springer Berlin / Heidelberg.

[ISO, 2005] ISO (2005). *Requirements for an Electronic Health Record Reference Architecture - ISO 18308.* International Organization for Standardization.

[Juric et al., 2007] Juric, M. B., Loganathan, R., Sarang, P., and Jennings, F. (2007). *SOA Approach to Integration - XML, Web services, ESB, and BPEL in real-world SOA projects*. PACKT Publishing.

[Kalra, 2006] Kalra, D. (2006). Electronic health record standards. *Yearbook of Medical Informatics*, pages 136–144.

[Kirsh, 2008] Kirsh, W. (2008). *Encyclopedia of Public Health*, volume 1. Springer Science.

[Lippeveld et al., 2000] Lippeveld, T., Sauerborn, R., and Bodart, C. (2000). *Design and Implementation of Health Information Systems*. World Health Organization, Geneva.

[Machado et al., 2006] Machado, J., Abelha, A., Neves, J., and Santos, M. (2006). Ambient intelligence in medicine. In *Proceedings of the IEEE-Biocas 2006 - Biomedical Circuits and Systems Conference*. Imperial College.

[Machado et al., 2010] Machado, J., Miranda, M., Gonçalves, P., Abelha, A., Neves, J., and Marques, J. (2010). Aidatrace: interoperation platform for active monitoring in healthcare environments. *8th International Industrial Simulation Conference 2010, ISC 2010*, pages 67–71.

[Marakas, 1998] Marakas, G. (1998). *Decision Support Systems: In the 21st Century*. Prentice Hall.

[Miranda et al., 2008] Miranda, M., Abelha, A., Santos, M., Machado, J., and Neves, J. (2008). A group decision support system for staging of cancer. In *Proceedings of the 1st International Conference on Electronic Healthcare in The 21st Century*, pages 114–121. Springer Berlin Heidelberg.

[O'Brien, 2009] O'Brien, R. (2009). *Systems Integration Made Friendly - An Overview for IT Managers and Developers*. Narrabeen, N.S.W.

[PentahoCommunity, 2014] PentahoCommunity (2014). Welcome to the pentaho community. http://community.pentaho.com. Acedido em Novembro de 2014.

[Rigor et al., 2008] Rigor, H., Machado, J., Abelha, A., Neves, J., and Alberto, C. (2008). A web-based system to reduce the nosocomial infection impact in healtcare units. In *WEBIST (1)*, pages 264–268.

[Santos and Azevedo, 2005] Santos, M. and Azevedo, C. (2005). *Data Mining, Descoberta de Conhecimento em Bases de Dados*. FCA editor.

[Shallahamer, 2007] Shallahamer, C. (2007). *Forecasting Oracle Performance*. Apress, Berkeley.

[Shohoud, 2002] Shohoud, Y. (2002). *Real World XML WEB Services - For VB and VB .NET Developers*. Addison-Wesley.

[Silva, 2012] Silva, P. (2012). Monitorização e prevenção de falhas em bases de dados hospitalares. Master's thesis, Universidade do Minho.

[Smith and Ceusters, 2007] Smith, B. and Ceusters, W. (2007). Hl7 rim: An incoherent standard. *Studies in Health Technology and Informatics*, 124:133–138.

[Sousa, 2002] Sousa, A. (2002). *Bases de dados Web e XML*. FCA Editora de Informática.

[Subbe et al., 2001] Subbe, C., Kruger, M., Rutherford, P., and Gemmel, L. (2001). Validation of a modified early warning score in medical admissions. *QJM*, 94(10):521–526.

[Sujansky et al., 2009] Sujansky, W. V., Overhage, J. M., Chang, S., Frohlich, J., and Faus, S. A. (2009). The development of a highly constrained health level 7 implementation guide to facilitate electronic laboratory reporting to ambulatory electronic health record systems. *Journal of the American Medical Informatics Association*, 16(3).

[Sullivan, 1992] Sullivan, M. (1992). System support for software fault tolerance in highly available database management systems. Technical report, University of California, Berkeley.

[Tange et al., 2001] Tange, A., Hasman, A., Robbe, P. V., and Schouten, H. (2001). Medical narratives in electronic medical records. *International Journal of Medical Informatics*, 46:7–28.

[Tereso and Bernardino, 2011] Tereso, M. and Bernardino, J. (2011). Open source business intelligence tools for smes. In *Information Systems and Technologies (CISTI), 2011 6th Iberian Conference*, pages 1–4.

[Thomsen and Pedersen, 2005] Thomsen, C. and Pedersen, T. (2005). A survey of open source tools for business intelligence. In *Data Warehousing and Knowledge Discovery, ser. Lecture Notes in Computer Science, A. Tjoa and J. Trujillo, Eds*, volume 3589, pages 74–84.

[Turban, 2004] Turban, E. (2004). *Decision support and expert systems: management support systems*. Prentice Hall PTR Upper Saddle River, 2nd edition edition.

[Turham et al., 2008] Turham, K., Kurt, B., and Uzun, E. (2008). Salik-net project in turkey and hl7 v3 implementation. *World Academy of Science, Engineering and Technology*, 34:317–321.

[Weed, 2008] Weed, L. (2008). Medical records, medical education, and patient care. *Irish Journal of Medical Science*, 39(6):271–282.

[Wei, 2012] Wei, X. (2012). Communications solutions for heterogeneous systems based on soa. *Physics Procedia*, 25:1738–1742.

[WHO, 2014] WHO (2014). International classification of diseases (icd) - world health organization. http://www.who.int/ classifications/icd/en/. Acedido em Julho de 2014.

[Yi, 2006] Yi, O. (2006). Highly available database systems. In *Integrated information systems and databases seminar*.

[Zaleski, 2009] Zaleski, J. (2009). Integrating device data into the electronic medical record: A developer's guide to design and a practitioner's guide to application, 1st edition. Technical report, Wiley-VCH.